

Uniwersytet Warszawski  
Wydział Nauk Ekonomicznych

Rafał Woźniak

# Ekonometryczne modelowanie popytu na energię elektryczną za pomocą danych wysokiej częstotliwości

Praca doktorska

Praca wykonana pod kierunkiem  
dr. hab. Ryszarda Kokoszczyńskiego, prof. UW  
z Zakładu Finansów Ilościowych  
WNE UW

Warszawa, czerwiec 2015

## **Podziękowania**

Dziękuję prof. Ryszardowi Kokoszczyńskiemu za pomoc w nadaniu ostatecznej struktury pracy oraz za cenne uwagi. Dłużny jestem podziękowania dr Natalii Nehrebeckiej za cierpliwe odpowiedzi na moje niekończące się pytania. Marcinowi Sienickiemu dziękuję za owocną dyskusję na temat zmiennych użytych w modelowaniu wskaźników koniunktury. Pani Annie Włostowskiej i Krystynie Karpiuk dziękuję za zrozumienie. Słowa podziękowania należą się Wojciechowi Broniarkowi za bycie pierwszym krytycznym czytelnikiem opisów. Szczególnie dziękuję Marcie za motywację.

# Spis treści

<b>Wstęp</b>	<b>5</b>
<b>1 Analiza jednozmiennowa</b>	<b>9</b>
1.1 Opis danych . . . . .	9
1.2 Analiza spektralna . . . . .	10
1.3 Stacjonarność i sezonowość . . . . .	14
1.4 Analiza komponentu trendu . . . . .	17
1.4.1 Zmiana czasu i brakujące obserwacje . . . . .	21
<b>2 Analiza sezonowości</b>	<b>23</b>
2.1 Model Holta-Wintersa . . . . .	24
2.2 Model Taylora . . . . .	26
2.3 Model Hyndmana i innych . . . . .	27
2.4 Modele ze stochastyczną i deterministyczną sezonowością . . . . .	29
2.4.1 Szczegółowa specyfikacja modelu . . . . .	31
2.4.2 Modele z zero-jedynkową sezonowością . . . . .	36
2.4.3 Modele z trygonometryczną sezonowością . . . . .	38
<b>3 Wpływ cyklu koniunkturalnego</b>	<b>41</b>
3.1 Równoczesny i wyprzedzający wskaźnik koniunktury . . . . .	42
3.2 Miesięczne PKB . . . . .	49
3.3 Dekompozycja szeregów energii elektrycznej . . . . .	55
3.3.1 Dane roczne . . . . .	56
3.3.2 Dane miesięczne . . . . .	57
3.3.3 Dane z łamaną częstotliwością . . . . .	58
3.4 Korelacja i przyczynowość . . . . .	61
3.5 Zmienne objaśniające . . . . .	64
<b>4 Zmienne egzogeniczne</b>	<b>67</b>
4.1 Zmienne pogodowe . . . . .	68
4.2 Analiza miesięczna . . . . .	70
4.3 Analiza danych dziennych . . . . .	75
4.4 Nieliniowe zależności . . . . .	77

4.5	Wpływ dni wolnych od pracy . . . . .	84
<b>5</b>	<b>Prognozowanie popytu na energię elektryczną</b>	<b>89</b>
5.1	Projekt eksperymentu . . . . .	91
5.2	Metoda dnia podobnego . . . . .	94
5.3	Model regresji . . . . .	97
5.4	Model autoregresyjny . . . . .	104
5.5	Model splajnów . . . . .	106
5.5.1	Model splajnów kubicznych . . . . .	106
5.5.2	Regresja splajnami sześciennymi . . . . .	109
5.5.3	Okresowe splajny sześcienne . . . . .	109
5.5.4	Specyfikacja modelu i wyniki . . . . .	111
5.6	Prognozowanie krótkookresowe . . . . .	114
5.7	Prognozowanie średniookresowe . . . . .	115
	<b>Zakończenie</b>	<b>121</b>
	<b>Bibliografia</b>	<b>123</b>
	<b>Aneks</b>	<b>135</b>

# Wstęp

Niewątpliwie energia elektryczna we współczesnym świecie odgrywa ogromną rolę. Specyfika elektryczności, a w zasadzie brak możliwości jej ekonomicznego magazynowania, sprawia, że analizowanie popytu na energię elektryczną jest niezmiernie ważne dla wszystkich podmiotów z rynku energii. Na rynku energii elektrycznej w Polsce, od momentu jego liberalizacji 1 lipca 2007, wyróżnia się wytwórców energii, pośredników, firmy zajmujące się przesyłem energii oraz odbiorców. Dysponowanie idealnymi prognozami zapotrzebowania na energię elektryczną mogłoby pozwolić każdemu uczestnikowi tego rynku na ekonomiczne gospodarowanie, ponieważ każde niedostateczne zakontraktowanie energii lub przeszacowanie zapotrzebowania generuje koszty. W centrum zainteresowania są wysokoczęstotliwościowe charakterystyki rynku energii, a dane 15-minutowe wydają się być idealne do tego typu analiz.

W niniejszej pracy sprawdzone zostały dwie główne hipotezy badawcze. Pierwsza z nich dotyczyła istnienia nieliniowego wpływu temperatury powietrza na zapotrzebowanie na energię elektryczną w obserwowanych danych 15-minutowych. Drugą zasadniczą hipotezą badawczą jest istotność pozytywnego wpływu cyklu koniunkturalnego na analizowane zapotrzebowanie na moce przesyłowe krajowych sieci elektroenergetycznych czyli sensowność wykorzystania prognoz zagregowanego stanu gospodarki do sporządzania prognoz średniookresowych. Wybór tych dwóch hipotez uargumentowany jest chęcią formalnego sprawdzenia postaci i pokazania nieliniowej zależności między temperaturą a zapotrzebowaniem na energię, natomiast druga, związana z cyklem koniunkturalnym, wydaje się niezwykle ciekawa dla autora.

Poza wymienionymi hipotezami głównymi wartymi rozważenia wydają się być jeszcze dodatkowe kwestie. Pierwszą hipotezą poboczną jest istnienie w obserwowanych danych kwadransowych opisujących zapotrzebowanie na moce przesyłowe potrójnej struktury sezonowej związanej z wahaniami rocznymi, tygodniowymi oraz fluktuacjami charakterystycznymi dla poszczególnych godzin. Drugim wartym weryfikacji stwierdzeniem jest to o istnieniu w rozważanych danych tendencji wzrostowej - trendu. Trzecim dodatkowym zagadnieniem rozważanym w pracy jest weryfikacja przypuszczenia o tym, że metoda prognozowania oparta o modele regresji liniowej dla poszczególnych kwadransów powinna okazać się najlepsza dla polskich danych.

Hipotezy główne i dodatkowe wydają się być ekonomicznie intuicyjne. Obserwowany wzrost gospodarczy Polski w okresie 2002-2014<sup>1</sup> przekładać się musi na wzrost zapotrzebowania na energię elektryczną. Oczekiwanie skomplikowanej struktury wahań okresowych zdaje się być bezdyskusyjne dla każdego analityka i uczestnika rynku energii. Cykle dni i nocy, tygodniowe struktury dni pracy i weekendów oraz serie pór roku, czyli zmiennej długości dnia, uzasadniają postawienie hipotezy o potrójnej strukturze wahań sezonowych. Hipoteza o przekładaniu się koniunktury na potrzeby elektroenergetyczne gospodarujących podmiotów wydaje się mieć uzasadnienie w teorii mikroekonomii. Przedsiębiorca lub firma planuje produkcję w zależności od bieżącego i spodziewanego popytu. Słabszy zagregowany popyt związany jest z okresami gorszej koniunktury, więc rozsądnym wydaje się oczekiwanie przekładania się sytuacji gospodarczej na energię elektryczną. W niektórych przypadkach wydawać się może, że zależność ta ma przeciwny zwrot, gdyż proces produkcyjny wyprzedza moment sprzedaży. Dokładnie z tego samego powodu oczekiwać można logiczności wykorzystania prognoz sytuacji ekonomicznej do wyznaczania przyszłych potrzeb energetycznych podmiotów. Zdecydowana większość badań wpływu cyklu koniunkturalnego na popyt na energię elektryczną wykorzystuje dane roczne i kwartalne. Wartość dodana niniejszej pracy polega także na analizowaniu wspomnianej relacji dla danych o wyższej, miesięcznej częstotliwości.

W niniejszej pracy podstawą analizy są dane 15-minutowego zapotrzebowania na moce przesyłowe krajowych sieci elektroenergetycznych (KSE). Informacje te bez wątpienia można określić mianem wysokoczęstotliwościowych. Dane są dostępne dla całej sieci, więc poziomem zagregowania jest cały kraj. Zamiennie będą używane pojęcia zapotrzebowania na moce przesyłowe KSE, popytu na energię i popytu na moce przesyłowe. Definitywnie nie są to tożsame pojęcia, jednak ze względu na poziom zagregowania danych (cały kraj) pojęcia te mogą być postrzegane jako tożsame. Pomijana jest tutaj kwestia eksportu i importu energii, ponieważ w wykonanych analizach nie dokonywane było rozróżnienie pochodzenia energii w sieci.

Praca jest skonstruowana tak, że kolejno opisywane są poszczególne składowe szeregu 15-minutowego zapotrzebowania na moce przesyłowe KSE i weryfikowane kolejne hipotezy. Pierwszy rozdział zawiera krótki opis analizowanej zmiennej oraz jej podstawowe charakterystyki. Główna część tego rozdziału poświęcona jest sprawdzeniu istnienia tendencji wzrostowej oraz analizom stacjonarności szeregu. Dla tych drugich zaproponowana została specjalna wersja testu stacjonarności sezonowych danych oraz zaprezentowane zostały jego stabilizowane wartości krytyczne. Wykorzystane zostało ponadto podstawowe narzędzie ekstrakcji komponentu trendu z szeregu oraz przedstawione zostały jego parametry, jakie powinny zostać zastosowane w przypadku danych 15-minutowych.

W rozdziale drugim zaprezentowane zostały analizy struktury sezonowej szeregu zapo-

---

<sup>1</sup>Silna tendencja wzrostowa w polskim produkcie krajowym brutto, związana z transformacją gospodarki, rozpoczęła się w połowie lat 90', jednak ze względu na zakres czasowy analizowanych danych przytoczony jest krótszy okres.

trzebowania na moce przesyłowe krajowych sieci elektroenergetycznych. Nie tylko sprawdzone zostało istnienie wielorakiej struktury wahań okresowych, ale także przeanalizowane zostały wersje tych wahań. Wychodząc od podstawowego modelu wygładzania wykładniczego zaprezentowane zostały jego kolejne rozszerzone wersje wyestymowane dla polskich danych. Zaprezentowane zostało dodatkowe podejście, które, jak się wydaje, nie zostało jeszcze wykorzystane w literaturze.

Rozdział trzeci poświęcony jest analizom cyklu koniunkturalnego i podstawowym sposobom sprawdzenia jego przekładania się na popyt na energię elektryczną. Ze względu na to, że w literaturze polskiej nie istnieje wiele opracowań zależności między zagregowanym stanem gospodarki a popytem na energię elektryczną w sieci zaprezentowane zostały powody, dla których istotne jest rozważanie tej kwestii oraz uwzględnienie jej w procesie modelowania i przede wszystkim prognozowania. Zaprezentowany został sposób oszacowania indeksu przybliżającego cykl koniunkturalny oraz jego wersja pozwalająca prognozować te wartości. Rozdział zawiera także podrozdział poświęcony propozycji szacowania szeregu miesięcznego produktu krajowego brutto, który mógłby najdokładniej odzwierciedlać sytuację w gospodarce.

Wykorzystanie zmiennych egzogenicznych w modelowaniu zapotrzebowania na energię elektryczną jest tematyką rozdziału czwartego. Zaprezentowane i wykorzystane zostały stochastyczne i deterministyczne zmienne objaśniające, czyli zmienne pogodowe i dni wolne od pracy. Ponownie, ze względu na nieliczność badań tego typu dla Polski, wykonana została szeroka analiza dla częstotliwości miesięcznych, tygodniowych i dziennych. Sprawdzone zostało istnienie nieliniowej zależności między temperaturą a energią za pomocą modeli parametrycznych i nieparametrycznych.

Rozdział piąty zawiera eksperymenty, w których porównane zostały sposoby prognozowania zapotrzebowania na energię. Rozważone zostały dwa horyzonty prognoz: 3-dniowy i 3-miesięczny. Wybór okresu 3-dniowego jest uargumentowany ogromnym zainteresowaniem teoretyków i praktyków prognozami w tym horyzoncie, natomiast okres 3-miesięczny umożliwia weryfikację hipotezy o użyteczności informacji o cyklu koniunkturalnym do prognozowania średniookresowego. Wykorzystane zostały modele uznane w literaturze za najdokładniejsze dla najczęściej opisywanego w literaturze rynku energii elektrycznej w Stanach Zjednoczonych.

Kończące pracę zakończenie zawiera rekapitulację najważniejszych wniosków i wyniki weryfikacji hipotez. Przedstawione zostały także kwestie wymagające dalszych prac i analiz w tej tematyce oraz potencjalne rozszerzenia zaprezentowanych metod.

Praca ma charakter akademicki z pewnymi elementami praktycznymi. Praktycy mogą mieć wypracowane własne metody i sposoby, które pozwalają uzyskać bardzo dobre rezultaty. Osoby zajmujące się zawodowo prognozowaniem popytu na energię elektryczną mogą mieć dostęp do komputerów z większymi możliwościami obliczeniowymi, dzięki czemu

mogą analizować bardziej złożone problemy. Dokładne prognozy można uzyskać na wiele sposobów, a przedstawione zostały te atrakcyjne i najbliższe autorowi.



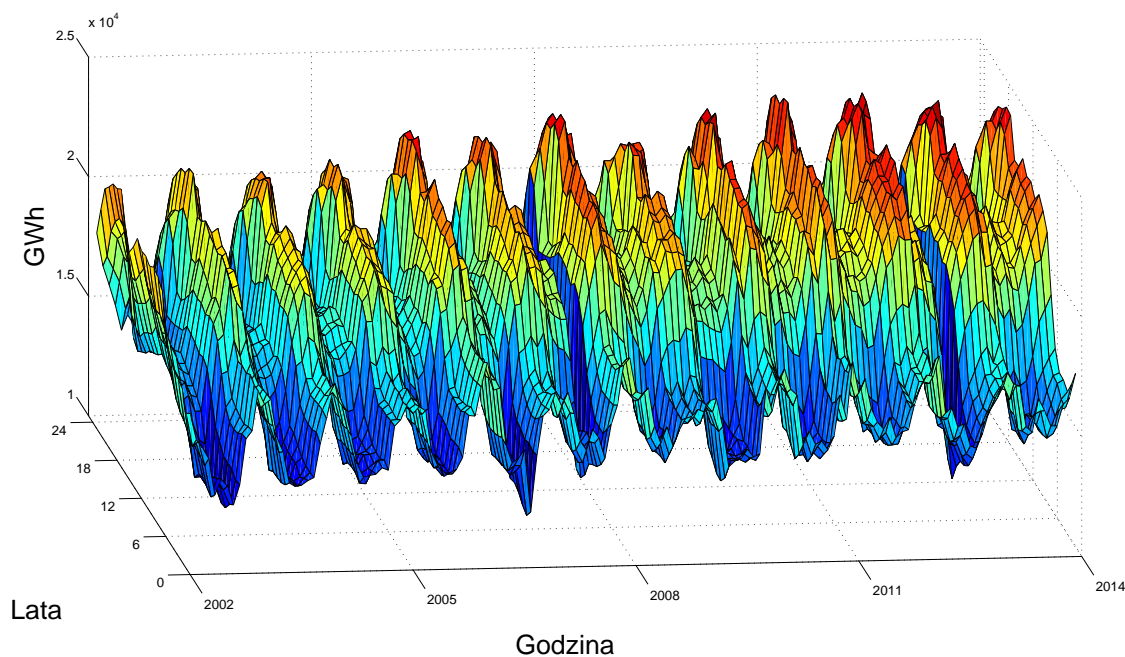
# Rozdział 1

## Analiza jednozmiennowa

### 1.1 Opis danych

W niniejszej pracy przedmiotem analiz jest szereg czasowy zapotrzebowania na moce przesyłowe krajowego systemu elektroenergetycznego. Dane dostępne są na stronie internetowej Polskich Sieci Elektroenergetycznych<sup>1</sup> od pierwszego kwadransu 2002 roku i

Rysunek 1.1: Zapotrzebowanie na moce przesyłowe KSE w trzecią środę miesiąca



*Źródło:* Opracowanie własne.

na bieżąco są rozszerzane. Dane próbkowane są co kwadrans, dzięki czemu mogą być nazywane wysokoczęstotliwościowymi. Wybrane do analizy zostały obserwacje od pierwszej

<sup>1</sup><http://www.pse.pl/index.php?dzid=77>

dostępnej za 1 stycznia 2002 r. do 10 października 2013 roku. Dane są danymi wysokoczęstotliwościowymi, ponieważ są próbkowane co kwadrans. W związku z tym, w próbie znajdują się 412892 obserwacje.

Analiza graficzna szeregu, przedstawionego na rysunku 1.1, wskazuje na występowanie tendencji wzrostowej w analizowanym 12-letnim okresie. Ze względu na przejrzystość na rysunku 1.1 zaprezentowane zostały tylko dane dla trzecich śród każdego miesiąca. Uzasadnione może być to ponadto tym, że zgodnie z nieformalną regułą postępowania zapewnić ma to lepszą porównywalność poprzez usunięcie części efektów sezonowych zaburzających wnioski. Zauważalne jest także zróżnicowanie wartości dla poszczególnych godzin oraz cykle większego i mniejszego zapotrzebowania zimą i latem. Wydaje się także, że istotny spadek miał miejsce w 2009 roku. Powyższe obserwacje zostały zweryfikowane w niniejszym, drugim i trzecim rozdziale.

Zapotrzebowanie na moce przesyłowe w krajowym systemie elektroenergetycznym jest zmienną stanu, w przeciwieństwie do wielu kategorii makroekonomicznych, które są przepływami. Oznacza to, że szereg ten musi być analizowany podobnie jak temperatura, stan wody w rzece czy kurs walutowy. Zmiana częstotliwości na niższą nie odbywa się, jak dla przepływów, poprzez sumowanie, ale poprzez wybranie jednej z obserwacji z analizowanego okresu, na przykład średniego poziomu lub wartości na koniec okresu.

W niniejszej pracy zajmuję się modelowaniem zapotrzebowania na moce przesyłowe krajowego systemu energetycznego. Niezwykle ważne dla wielu uczestników rynku energii jest także modelowanie i prognozowanie cen energii elektrycznej. Obszerną analizę metod prognozowania cen energii przedstawił Weron<sup>2</sup>.

## 1.2 Analiza spektralna

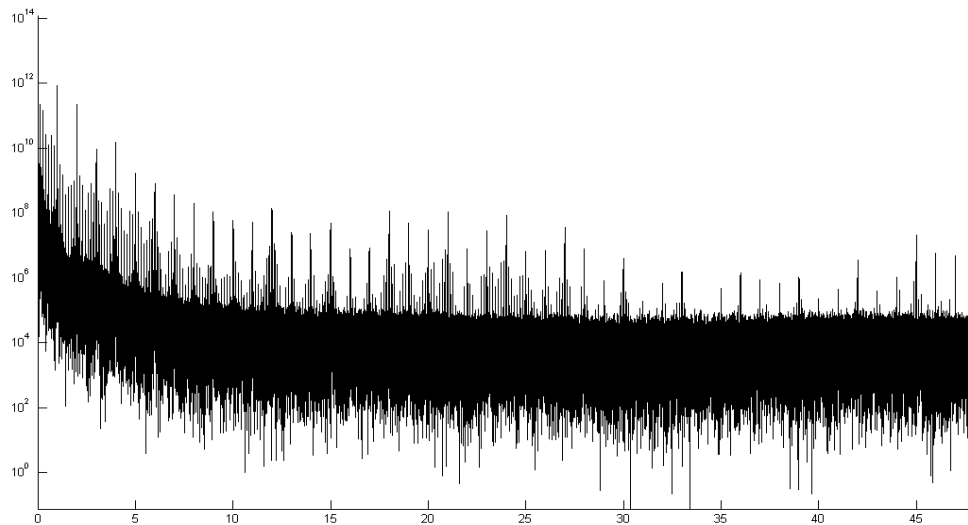
Analiza częstotliwościowa pozwala zbadać występowanie wahań o poszczególnych częstościach w szeregu czasowym. Dane o zapotrzebowaniu na moce krajowej sieci elektroenergetycznej potraktowane zostały w poniższych analizach jako dane dzienne, próbkowane 96 razy na każdy dzień, czyli co kwadrans. Alternatywnie uznać można, że okres wynosi tydzień, a próbkowanie równa się 672 razem w tygodniu. Wybór okresu nie wpływa na wnioski, a jedynie na sposób interpretacji wyników. W typowych analizach zmiennych miesięcznych lub kwartalnych za okres przyjmuje się odpowiednio 12 lub 4 obserwacje, natomiast w przypadku danych o złożonej strukturze sezonowej wybór ten nie jest jednoznaczny.

W widmie szeregu, przeskalowanego do skali logarytmicznej, widoczne są składowe odpowiadające wahaniom długookresowym. Potwierdza to tezę o niestacjonarności szeregu.

---

<sup>2</sup>Weron R., 2014, *Electricity Price Forecasting: A Review of the State-of-the-Art with a Look into the Future*, International Journal of Forecasting, 30, s. 1030-1081.

Rysunek 1.2: Widmo oryginalnego szeregu - skala logarytmiczna



Źródło: Opracowanie własne.

Argumenty widma rozważanego szeregu przebiegają przedział  $[0, 48]$ , gdzie 48 jest częstotliwością Nyquista. Transformacji Fouriera poddawane są szeregi stacjonarne, w związku z czym z danych usunięty został deterministyczny trend liniowy. Uzasadnione jest to tym, że takie działanie nie zmienia charakterystyki częstotliwościowej szeregu poza nieskończonymi częstotliwościami.

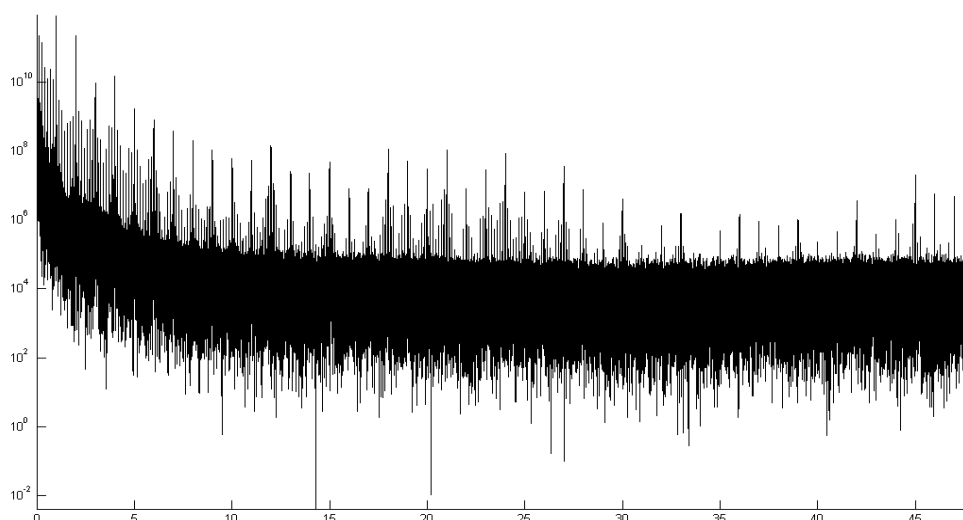
Pik w widmie oszacowanym dla częstości 1 odpowiadającej wahaniom powtarzającym się 1 raz na 96 kwadransów świadczy o występowaniu sezonowości przypisywanej poszczególnym kwadransom, czyli sezonowości dziennej. Piki powtarzające się dla kolejnych całkowitoliczbowych częstości odpowiadają harmonicznym sezonowości dziennej i wynikają z własności transformaty Fouriera.

Zawężając analizę gęstości spektralnej procesu do przedziału  $[0, 1]$ , wnioskowanie zostaje ograniczona do wahań rzadszych niż 1 raz na dzień, co prezentuje rysunek 1.4. Zauważyć można lokalne maksimum mocy dla argumentu 0,1428, co jest przybliżeniem  $\frac{1}{7}$ , czyli wahaniom powielającym się co 7 dni i związane z sezonowością tygodniową. Piki dla częstości  $\frac{2}{7}, \frac{3}{7}, \dots, \frac{6}{7}$  są harmonicznymi tej sezonowości. Rysunek 1.5 przedstawia widmo ograniczone do jeszcze węższego przedziału. Lokalne maksimum funkcji mocy dla częstości 0,00279, oraz jej harmonicznymi, związane jest z wahaniami powielającymi się co około 365,25 dnia. Poprawka o 0,25 wynika z kwestii roku przestępnego pojawiającego się w próbkę raz na 4 lata. Podejście to jest stosowane przez autorów zajmujących się tematyką sezonowości w danych dziennych, na przykład Dordonnat i inni<sup>3</sup>.

Warto zauważyć także wysokie wartości widma dla niskich częstości. Wydaje się, że -

<sup>3</sup>Dordonnat V., Koopman S.J., Ooms M., Dessertaine A., 2008, *An Hourly Periodic State Space Model for Modeling French National Electricity Load*, International Journal of Forecasting, 24, 588-587.

Rysunek 1.3: Widmo szeregu z odjętym trendem liniowym - skala logarytmiczna



Źródło: Opracowanie własne.

oprócz trendu - w danych występować może składnik cykliczny, odpowiadający wahaniom średniookresowym. Na rysunku 1.4 świadczyć o tym może lokalne maksimum mocy dla częstości 0,001163, czyli dla wahań co 2,35 roku. Zgodnie z najczęściej stosowaną definicją cyklu koniunkturalnego Burnsa i Mitchella<sup>4</sup> za wahania odpowiadające cyklowi koniunkturalnemu uznać należy te, które powtarzają się w zakresie od półtora roku do 8 lat. Uzyskana wartość wpada w definicyjny przedział.

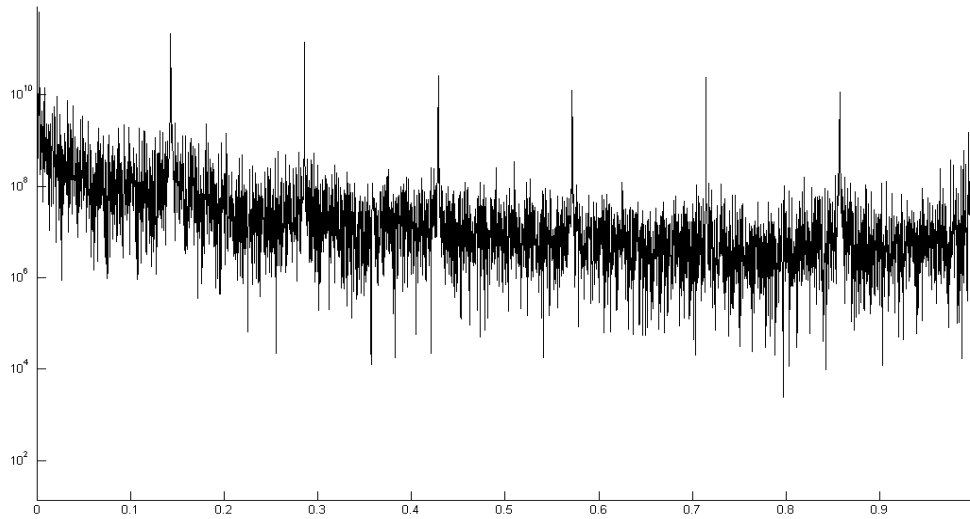
Obecność zaburzeń w obserwowanym szeregu może wpływać na uzyskiwane widmo mocy. Powoduje to problem tzw. wycieku mocy lub ucieczki częstotliwości (*spectral leakage*), który skutkuje przeniesiem części mocy z właściwej częstotliwości np.:  $\omega = \omega_0$  do innej  $\omega = \omega_a$ . Rozwiązaniem tego problemu jest zastosowanie okienkowania sygnału. W świetle twierdzenia o splocie przemnożenie oryginalnego sygnału przez okienko jest równoznaczne w domenie częstotliwości ze splotem jego funkcji mocy z okienkiem. Problem wycieku mocy został opisany między innymi przez Harrisa<sup>5</sup>, który też porównał różne okienka.

Sygnał został zokienkowany trzema różnymi, często wykorzystywanymi, oknami Blackmana-Harrisa, Hamminga oraz Tuckeya. Uzyskana w ten sposób gęstość widmowa sygnału jest bardzo podobna w każdym z tych przypadków do oryginalnego sygnału. Rysunek 1.6 prezentuje część widm mocy uzyskanych z zastosowania trzech wspomnianych okien. W przypadku okien Blackmana-Harrisa i Hamminga obserwowany jest pik dla częstości 0,001163, odpowiadającej wahaniom powtarzającym się co 2,35, co daje podstawę do dalszej analizy obecności wahań średniookresowych w szeregu zapotrzebowania mocy krajo-

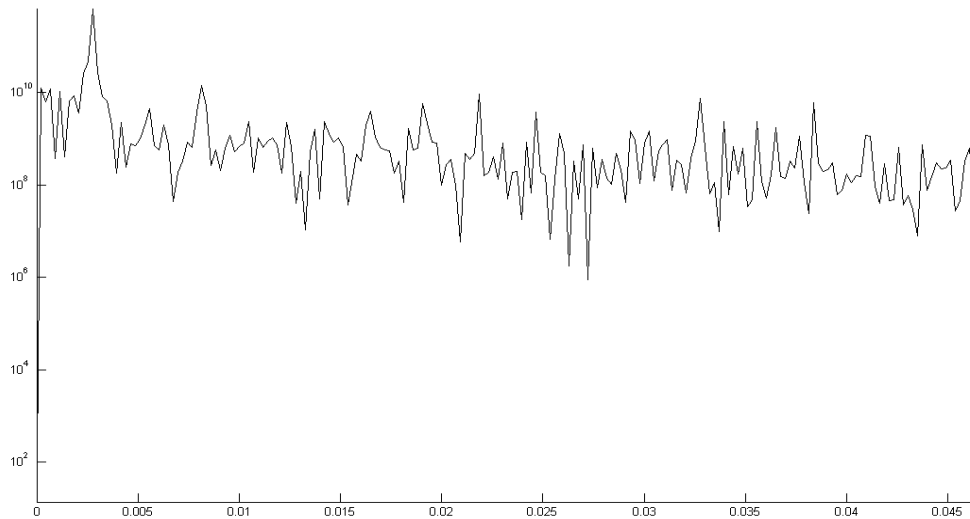
<sup>4</sup>Burns A.F., Mitchell W.C., 1938, *Statistical Indicators of Cyclical Revival*, National Bureau of Economic Research, Nowy Jork.

<sup>5</sup>Harris, 1976, *Windows, Harmonic Analysis, and the Discrete Fourier Transform*, Naval Undersea Center, San Diego.

Rysunek 1.4: Widmo szeregu z odjętym trendem liniowym ograniczone do przedziału  $[0, 1]$  - skala logarytmiczna

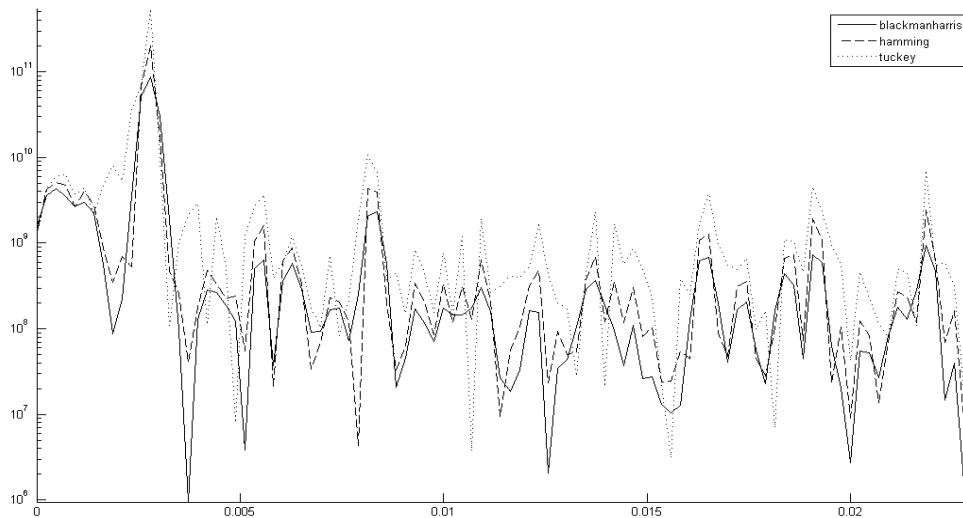


Rysunek 1.5: Widmo szeregu z odjętym trendem liniowym ograniczone do przedziału  $[0, 0,463]$  - skala logarytmiczna



*Źródło:* Opracowanie własne.

Rysunek 1.6: Część okienkowanego widma szeregu pozbawionego trendu deterministycznego w skali liniowej



*Źródło:* Opracowanie własne.

wych sieci elektroenergetycznych. Zgodnie z artykułem Harrisa<sup>6</sup> dla zastosowanych okien konkurencyjnymi właściwościami charakteryzują się okna Kaisera-Bessela oraz Dolpha-Chebysheva. Fragment gęstości widmowej wyznaczonej za pomocą tych dwóch okien przedstawiony został w aneksie na rysunku 1.6. Wnioski są analogiczne do przedstawionych powyżej i uzyskanych za pomocą okien Blackmana-Harrisa, Hamminga oraz Tuckeya.

### 1.3 Stacjonarność i sezonowość

Formalne przetestowanie występowania trzech rodzajów sezonowości - rocznej, tygodniowej i dziennej - w rozważanych danych wymaga oszacowania modelu i przetestowania hipotezy łącznej. W związku z tym, oszacowane zostały dwa modele regresji dla zlogarytmowanej zmiennej  $y_t$  oraz dla jej wersji z usuniętym trendem liniowym  $y_t^{bt}$ , a ich wyniki przedstawione są w tabelicy 1.1 w kolumnach (1)-(2). Na ich podstawie należy odrzucić hipotezę o łącznej nieistotności trzech rodzajów sezonowości. Ewentualnym zastrzeżeniem do tych modeli może być skromna wartość p-value testów autokorelacji. Modele (1)-(2), jak i (3)-(4), zostały oszacowane w programie Stata.

---

<sup>6</sup>Tamże.

Tablica 1.1: Modele testujące stacjonarność i sezonowość danych

zmienna	(1) $\ln(y_t)$	(2) $\ln(y_t^{bt})$	(3) $\Delta y_t$	(4) $\Delta y_t$
<i>stacjonarność</i>				
$\delta$	-	-	-53,51	-24,59
<i>stała</i>				
<i>opóźnienia zmiennej zależnej</i>	tak 1;5	tak 1;5	tak 1/35	nie 1/35
<i>test braku sezonowości</i>				
statystyka testowa	348,61	348,18	-	-
p-value	0,000	0,000	-	-
<i>test braku autokorelacji</i>				
statystyka testowa	8,425	9,571	1,018	3,197
p-value	0,004	0,002	0,313	0,074

Źródło: Opracowanie własne.

Pozostałe kolumny tablicy 1.1 poświęcone są wynikom badania rzędu stacjonarności. Najczęściej stosowanym testem do testowania stopnia integracji jest test Dickey’a-Fullera<sup>7</sup>, w skrócie DF. Testowana jest hipoteza  $H_0 : \delta = 0$ , przeciwko alternatywie  $H_1 : \delta < 0$  w równaniu

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t. \quad (1.1)$$

Test nazywany jest też testem pierwiastka jednostkowego, ponieważ równanie (1.1) może zostać zapisane w postaci

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad (1.2)$$

gdzie  $\rho = (1 + \delta)$ , która jest równoważna 1.1. Odrzucenie hipotezy zerowej  $H_0$  na korzyść alternatywy oznacza, że szereg  $y_t$  jest zintegrowany w stopniu zero. Ze względu na obciążenie estymatora parametru  $\delta$  sposób weryfikacji założonej hipotezy jest odmienny od standardowo przyjętego w modelach regresji liniowej. Rozkład statystyki testowej  $t = \hat{\delta}/se(\hat{\delta})$  nie jest rozkładem t-Studenta, a jego tablice mogą być generowane za pomocą metody Monte Carlo. Tablice rozkładu statystyki testowej zostały przedstawione przykładowo przez Charemzę i Deadmana<sup>8</sup>. W przypadku sezonowego szeregu czasowego w równaniu (1.1) pojawiają się dodatkowe regresory odpowiadające sezonowości. Dla danych 15-minutowych z wahaniami sezonowymi równanie (1.1) przyjmie postać:

$$\Delta y_t = \delta y_{t-1} + \sum_{i=1}^{96} Q_{K,i,t} + \sum_{i=1}^6 Q_{D,i,t} + \sum_{i=1}^{11} Q_{M,i,t} + \epsilon_t, \quad (1.3)$$

<sup>7</sup>Dickey D.A., Fuller W.A., 1979, *Distribution of Estimators for Autoregressive Time Series With a Unit Root*, Journal of American Statistical Association, Vol. 74, No. 366, s. 427-431

<sup>8</sup>Charemza W., Deadman D. 1997, *Nowa ekonometria*, Polskie Wydawnictwo Ekonomiczne, s. 292

gdzie  $Q_{K,i,t}$ ,  $Q_{D,i,t}$ ,  $Q_{M,i,t}$  są zmiennymi zero-jedynkowymi odpowiadającymi odpowiednio poszczególnym kwadansom, dniom i miesiącom. Na podstawie modeli (3)-(4) wyciągnąć należy wniosek o zerowym stopniu zintegrowania szeregu. Konkluzja ta jest sprzeczna z analizą graficzną szeregu i przypuszczeniami. Prawdopodobną przyczyną takiego rezultatu może być występowanie autokorelacji wysokiego rzędu, występowanie także stochastycznej wersji sezonowości, stochastycznego trendu lub niska moc testu DF, a rozwiązaniem umożliwiającym doprecyzowanie wniosków może być bardziej skomplikowany test pierwiastka jednostkowego.

W związku z wymienionymi słabościami nie można poprzestać na tym podejściu należy wykorzystać inne narzędzia. Test zaproponowany przez Hylleberga i innych<sup>9</sup>, zwany dalej HEGY, jest testem pozwalającym zidentyfikować rodzaj sezonowości i jednocześnie zweryfikować hipotezę o istnieniu pierwiastka jednostkowego. Test HEGY dla danych 15-minutowych opiera się na modelu regresji postaci:

$$\Delta_{96}y_t = \sum_{i=1}^{96} a_i Q_{i,t} + \sum_{i=1}^{96} b_i Y_{i,t-1} + \sum_{i=1}^k c_i \Delta_{96}y_{t-i} + \epsilon_t, \quad (1.4)$$

gdzie  $k$  jest liczbą opóźnień zmiennej objaśnianej, zmienne  $Q_i$  są sezonowymi zmiennymi zero-jedynkowymi, natomiast  $Y_{i,t}$  są zmiennymi zdefiniowanymi w następujący sposób:

$$Y_{1,t} = \sum_{j=1}^{96} \cos(j \cdot 0) y_{t-j}, \quad (1.5)$$

$$Y_{2i,t} = \sum_{j=1}^{96} \cos(j2\pi \cdot (2i)/96) y_{t-j}, \quad (1.6)$$

$$Y_{2i+1,t} = \sum_{j=1}^{96} \sin(j2\pi \cdot (2i+1)/96) y_{t-j}, \quad (1.7)$$

dla  $i = 1, 2, \dots, 47$  oraz

$$Y_{96,t} = \sum_{j=1}^{96} \cos(j2\pi \cdot 96/96) y_{t-j}. \quad (1.8)$$

Zgodnie z interpretacją Charemzy i Deadmana<sup>10</sup>, gdy wszystkie współczynniki  $b_i$  są równe zeru, a współczynniki  $a_i$  jednakowe, to w obserwowanym szeregu  $y_t$  występuje tylko stochastyczna sezonowość. W przypadku różnych wartości  $a_i$  i przynajmniej jednego  $b_i$  różnego od zera obserwowana jest zarówno stochastyczna, jak i deterministyczna wersja wahań sezonowych. Poszczególne zmienne  $Y_i$  odpowiadają wahaniom z określonymi częstotliwościami. Tak na przykład  $Y_1$  odpowiada integracji szeregu, a obydwie  $Y_2$  oraz  $Y_3$  wahaniom sezonowym powtarzającym się co 96 okresów 15-minutowych.

<sup>9</sup>Hylleberg S., Engle R.F., Granger C.W.J., Yoo B.S., 1990, *Seasonal Integration and Cointegration*, Journal of Econometrics, Vol. 44(1-2), s. 215-238

<sup>10</sup>Charemza W., Deadman D. 1997, *Nowa ekonometria...*



Tablica 1.2: Statystyki testowe i wartości krytyczne testu HEGY

Hipoteza	$\Delta_{96}y_t$	Wartość krytyczna 5%
$H_0^1 : b_1 = 0$	28,625	-2,82
$H_0^2 : b_2 = b_3 = 0$	1654,452	3,37
$H_0^3 : \forall_i a_i = 0$	26278,304	3,43

Źródło: Opracowanie własne.

Z powodu braku programu umożliwiającego przeprowadzenie testu HEGY dla danych 15-minutowych konieczne było samodzielne oprogramowanie tego testu<sup>11</sup>. Tablica 1.2 prezentuje wyniki testu HEGY dla badania stacjonarności analizowanych danych 15-minutowych. Dodatnia wartość statystyki testowej w testowaniu  $H_0^1$  pozwala stwierdzić występowania pierwiastka jednostkowego, co potwierdza dodatkową hipotezę badawczą o występowaniu rosnącego trendu. Potwierdzić należy występowanie stochastycznej sezonowości ze względu na wartość statystyki testowej 1654,452 wykraczającej poza 95% przedział ufności. Co więcej, uznać należy także fakt istotności deterministycznej sezonowości w rozważanym szeregu.

Test HEGY został przeprowadzony raz jeszcze, z tą różnicą, że zmienną objaśnianą w równaniu (1.4) były różnice ze skorygowanym odjemnikiem  $\tilde{\Delta}_{96}y_t$ . Różnice ze skorygowanym odjemnikiem różnią się od klasycznych różnic rzędu 96 tym, że dla okresów związanych ze zmianą czasu odjemnikiem jest 95 lub 97 opóźnienie zmiennej. Korekta ta służyć ma sprawdzeniu czy zmiana czasu związana z wprowadzaniem czasu letniego nie zaburza wnioskowania. Wnioski z testu HEGY dla właściwego różnicowania zmiennej pozostają jednak niezmiennie.

## 1.4 Analiza komponentu trendu

Filtr Hodricka-Prescotta<sup>12</sup> został zaproponowany około 18 lat temu i obecnie jest jednym ze standardowych narzędzi ekonometrycznych. Narzędzie to służy do wyodrębniania z szeregu czasowego długookresowego trendu. Różnica między oryginalnym i wynikowym szeregiem stanowi wahania cykliczne i nieregularne.

Filtr Hodricka-Prescotta, dalej HP, jest liniowym, dwustronnym filtrem za pomocą którego można wyznaczyć wygładzony komponent  $\mu$  szeregu  $y$ . Szkielet filtru stanowi przedstawienie szeregu czasowego  $y_t$  w postaci sumy komponentu wzrostowego  $\mu_t$  oraz kompo-

<sup>11</sup>Stworzony w tym celu został skrypt "HEGY96.ado", który jest instalowalną procedurą dla pakietu Stata.

<sup>12</sup>Hodrick R.J., Prescott E.C., 1997, *Postwar U.S. Business Cycles: An Empirical Investigation*, Journal of Money, Credit and Banking, Vol. 29, No. 1, s. 1-16.

mentu zawierającego wahania średnio- i krótkookresowe  $c_t$ :

$$y_t = \mu_t + c_t \text{ dla } t = 1, 2, \dots, T. \quad (1.9)$$

Komponent  $c$  interpretowany jest jako odchylenia od  $\mu$ , i z założenia w długim okresie suma jego wartości równa się zeru. Na podstawie powyższego zapisać można następujący problem optymalizacyjny:

$$\min_{\mu_t} \left\{ \sum_{t=1}^T c_t^2 + \lambda \sum_{t=1}^T [(\mu_t - \mu_{t-1}) - (\mu_{t-1} - \mu_{t-2})]^2 \right\} \quad (1.10)$$

gdzie  $c_t = y_t - \mu_t$ . Parametr  $\lambda$  jest dodatnią stałą, nazywaną parametrem kary, gdyż jego rolą jest "karanie" za zbyt duże wahania w wygładzonym komponencie  $c_t$ . Wzrost wartości  $\lambda$  wpływa na stopień wygładzenia wynikowego szeregu. W granicy, gdy  $\lambda \rightarrow \infty$ , rozwiązanie zagadnienia (2) odpowiada dopasowaniu liniowego trendu czasowego z klasycznej regresji. Wartość parametru  $\lambda$  dobierana jest na podstawie częstotliwościowej reguły Ravna i Uhliga<sup>13</sup>, według której liczbę okresów w ciągu roku należy podzielić przez 4, podnieść do czwartej potęgi i przemnożyć przez 1600. Ravn i Uhlig proponują stosowanie czwartej potęgi, natomiast zastosowanie drugiej potęgi prowadzi do uzyskania wartości  $\lambda$  z oryginalnej pracy Hodricka i Prescottta. Zagadnienie 1.10 może być elegancko przedstawione w postaci modeli przestrzeni stanów. Zgodnie z Harvey<sup>14</sup> rozważmy model lokalnego liniowego trendu ze stochastycznym nachyleniem i deterministycznym poziomem:

$$y_t = \mu_t + \varepsilon_t, \quad (1.11)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1}, \quad (1.12)$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad (1.13)$$

gdzie  $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$  oraz  $\zeta_t \sim NID(0, \sigma_\zeta^2)$ .

Gdy stosunek wariancji błędu w równaniu pomiarowym (1.11) do wariancji błędu w równaniu stanu (1.13) zostanie ustalony na poziomie  $\lambda = \sigma_\varepsilon^2 / \sigma_\zeta^2$ , wtedy układ równań (1.11)-(1.13) jest zapisem filtru HP, gdzie  $\mu_\tau$  jest wygładzonym szeregiem. Konsekwencją niewłaściwego wyboru parametru  $\lambda$  może być otrzymanie nieprawdziwych wartości komponentu trendu. Filtr Hodricka-Prescottta nie jest narzędziem idealnym, jakkolwiek pozostaje jedną z podstawowych metod usuwania trendu. Inne podejście do analizy wahań długookresowych za pomocą filtru Hodricka-Prescottta zostało przedstawione przez Weronę i Zatora<sup>15</sup>, którzy wykorzystali analizę falkową do szeregu cen energii elektrycz-

<sup>13</sup>Ravn M.O., Uhlig H., 2002, *On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations*, The Review of Economics and Statistics, Vol. 84, No. 2, s. 371-375.

<sup>14</sup>Harvey A.C., 1989, *Forecasting, Structural Time Series Models...*, str. 45.

<sup>15</sup>Weron R., Zator M., 2015, *A Note on Using the Hodrick-Prescott Filter in Electricity Markets*, Energy Economics, 48(C), s. 1-6.

nej.

Weron i Zator<sup>16</sup> stwierdzili, że filtr Hodricka-Prescotta może być z powodzeniem stosowany do szeregów czasowych opisujących rynek energii elektrycznej. Wykorzystanie tego filtra wiąże się z potrzebą określenia odpowiedniej, do niestandardowej częstotliwości danych, wartości parametru  $\lambda$ . Poniżej prezentuję sposób wyznaczenia właściwej wartości  $\lambda$  poprzez analizę funkcji przenoszenia i wzmocnienia (*gain*) filtra. Funkcje te pozwalają zilustrować wpływ działania filtra na charakterystykę częstotliwościową filtrowanego szeregu. Funkcja przenoszenia częstotliwości lub transmitacji jest równa stosunkowi transformaty sygnału szeregu przefiltrowanego do transformaty sygnału filtrowanego. Funkcja wzmocnienia filtra jest równa modułowi funkcji przenoszenia<sup>17</sup>. Szczegółowe opisy funkcji generującej autokowariancję, funkcji przenoszenia i wzmocnienia znaleźć można w Hamilton<sup>18</sup> i Skrzypczyński<sup>19</sup>.

W celu wyznaczenia funkcji zysku z filtra Hodricka-Prescotta należy skorzystać z jego reprezentacji przestrzeni stanów (1.11)-(1.13) oraz z funkcji generującej autokowariancję dla stałej oraz elementu losowego. Niech  $L$  będzie operatorem opóźnień takim, że  $L^k y_t = y_{t-k}$ . Przekształcając równania (1.11)-(1.13) otrzymuje się

$$\mu_t = \frac{\zeta_t}{(1 - L)^2}.$$

Wobec powyższego funkcja generująca autokowariancję dla komponentu  $\mu_t$  ma postać

$$g_\mu(L) = \frac{1}{(1 - L)^2} \frac{1}{(1 - L^{-1})^2} \sigma_\zeta^2. \quad (1.14)$$

Stosując transformację Fouriera, podstawiając  $L = e^{-i\omega}$  oraz wykorzystując wzór Eulera  $e^{i\omega} = \cos(\omega) + i \sin(\omega)$ , równanie (1.14) może zostać zapisane w formie

$$g_\mu(\omega) = \frac{1}{(2 - 2 \cos(\omega))^2} \sigma_\zeta^2 \quad (1.15)$$

Funkcja zysku z filtra Hodricka-Prescotta określona jest wzorem

$$B^{HP}(\omega) = \frac{g_\mu(\omega)}{g_\epsilon(\omega) + g_\mu(\omega)} = \frac{1}{1 + \frac{\sigma_\epsilon^2}{\sigma_\zeta^2} (2 - 2 \cos(\omega))^2}. \quad (1.16)$$

Wartość  $\lambda$  wyliczona zgodnie z formułą zaproponowaną przez Hodricka i Prescottta dla danych 15-minutowych pozwala otrzymać filtr dolno-przepustowy, który przepuszcza zbyt wiele częstotliwości, oprócz pożądaných przepuszcza także wyższe niż 10-letnie. W związku

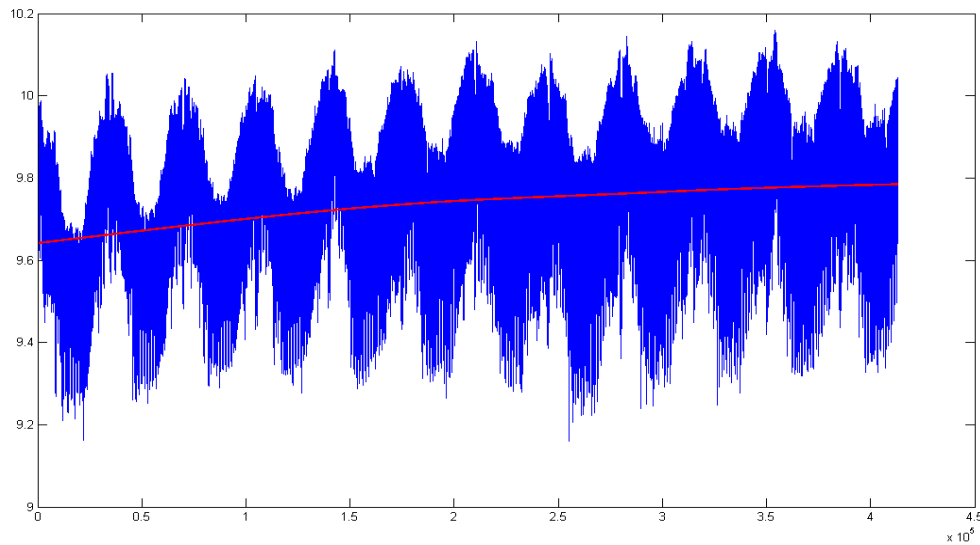
<sup>16</sup>Weron R., Zator M., 2015, *A Note on Using ...*, str. 5.

<sup>17</sup>Harvey A.C., Trimbur T.M., 2003, *General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series*, The Review of Economics and Statistics, 85(2), str. 245.

<sup>18</sup>Hamilton J.D., 1994, *Time Series Analysis*, Princeton University Press, Princeton.

<sup>19</sup>Skrzypczyński P., 2010, *Metody spektralne w analizie cyklu koniunkturalnego gospodarki Polskiej*, Materiały i Studia nr 252, Narodowy Bank Polski.

Rysunek 1.7: Szereg zapotrzebowania na moce przesyłowe i jego trend



Źródło: Opracowanie własne.

z powyższym należy zastosować regułę Ravna i Uhliga, w której

$$\lambda = 1600 \cdot \left(\frac{l}{4}\right)^4,$$

gdzie  $l$  jest liczbą okresów w roku. Identyczny wynik otrzymuje się stosując wzór przedstawiony przez Harvey'a i Trimbur<sup>20</sup>, gdzie

$$\lambda = \left[2 \sin\left(\frac{\omega_c}{2}\right)\right]^4,$$

gdzie  $\omega_c$  jest wybraną częstością odcięcia.

Estymacja komponentu trendu  $\mu$  została wykonana w programie MatLab na podstawie własnej implementacji modelu. W zastosowanym filtrze wartość  $l$  była równa  $l = 4 \cdot 24 \cdot 365 \cdot 10$ , co odpowiada  $\omega_c = \frac{2\pi}{4 \cdot 24 \cdot 365 \cdot 10}$ . Funkcja wiarygodności została zmaksymalizowana za pomocą algorytmu wyżarzania symulowanego, a uzyskana wartość funkcji log-wiarygodności równa była 161310,60. Oszacowana wartość wariancji składnika losowego  $\hat{\sigma}_{varepsilon}^2 = 0,026798$ , przy odchyleniu standardowym tego oszacowania  $4,0251 \cdot 10^{-9}$ . Statystyka istotności oszacowania jest argumentem potwierdzającym istnienie niestacjonarnego komponentu w danych o zapotrzebowaniu na moce krajowej sieci energetycznej. Dodatkowo, rysunek 2 potwierdza założoną hipotezę o występowaniu pierwiastka jednostkowego w analizowanym szeregu.

<sup>20</sup>Harvey A.C., Trimbur T.M., 2003, *General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series*, The Review of Economics and Statistics, 85(2), str 245.

### 1.4.1 Zmiana czasu i brakujące obserwacje

Zgodnie z Dyrektywą 2000/84/WE Parlamentu Europejskiego i Rady<sup>21</sup> z dnia 19 stycznia 2001 r. w sprawie ustaleń dotyczących czasu letniego w krajach należących do Unii Europejskiej, w każdą ostatnią niedzielę marca następuje zmiana czasu uniwersalnego o godzinę, czyli przejście z czasu zimowego na letni. W związku z tym uznaje się, że po godzinie 1:00 następuje godzina 3:00, w efekcie doba tego dnia trwa 23 godziny. Uzasadnione jest to lepszym dostosowaniem aktywności ludzi do godzin, w których jest najwięcej światła słonecznego. Zmiana czasu letniego na zimowy ma miejsce w ostatnią niedzielę października i polega przestawieniu zegarów o 60 minut do tyłu o godzinie 3:00.

Aries i Newsham<sup>22</sup> przytoczyli dane, które wskazują, że w krajach uprzemysłowionych konsumpcja elektryczności związana z oświetleniem odpowiada za 5-15% całkowitego zużycia energii elektrycznej<sup>23</sup>. Przywołany przez Aries i Newshama<sup>24</sup> raport Instytutu Polityki Fiskalnej w Indianie nie wskazał jednoznacznych wniosków dla tamtejszych danych, podczas gdy późniejsze badanie Kotchena i Grant<sup>25</sup> zaprzeczyło istnieniu oszczędności wynikających ze zmian czasu i wskazało na istotny wzrost zużycia energii tym wywoływany dla danych dla tego samego stanu. Badania przedstawione w Aries i Newsham<sup>26</sup> oraz przez Hilla i innych<sup>27</sup> wskazują na niejednoznaczność efektów wprowadzania czasu letniego. Racjonalność wprowadzania czasu letniego nie może być analizowana tylko i wyłącznie na podstawie statystyk zużycia energii elektrycznej. Aries i Newsham odnotowali, że dodatkowa godzina wieczorem przed zmierzchem w niektórych częściach Stanów Zjednoczonych przelożyła się na wzrost zużycia paliw poprzez zmianę zachowań ludzi. Wspomnieć tutaj także należy analizę Ferguson i innych<sup>28</sup> oraz raport Królewskiego Towarzystwa Zapobiegania Wypadkom przytoczony w Aries i Newsham<sup>29</sup> pokazujące pozytywną zależność między czasem letnim a zmniejszeniem liczby ofiar wypadków samochodowych i potrażeń pieszych i rowerzystów. Dyskusja na temat zmiany czasu jest bardzo szeroka i nie wskazuje na jednoznaczne wnioski, więc weryfikacja racjonalności czasu letniego w Polsce pozostanie tematyką dalszych prac badawczych.

Zagadnienie zmiany czasu nie było rozpatrywane w uprzednio przedstawionych modelach,

<sup>21</sup>Parlament Europejski i Rada Unii Europejskiej, 2001, *Dyrektywa 2000/84/WE Parlamentu Europejskiego i Rady z dnia 19 stycznia 2001 r. w sprawie ustaleń dotyczących czasu letniego*, Dziennik Urzędowy Unii Europejskiej, 12 t.2.

<sup>22</sup>Aries M.B.C., Newsham G.R., 2008, *Effect of Daylight Saving Time on Lighting Energy Use: A Literature Overview*, Energy Policy, 36, s. 1858-1866.

<sup>23</sup>Aries M.B.C., Newsham G.R., 2008, *Effect of Daylight Saving ...*, str. 1858

<sup>24</sup>Tamże, str. 1862

<sup>25</sup>Kotchen M.J., Grant L.E., 2008, *Does Daylight Saving Time Save Energy? Evidence from a Natural Experiment in Indiana*, NBER Working Paper 14429.

<sup>26</sup>Aries M.B.C., Newsham G.R., 2008, *Effect of Daylight Saving ...*, str. 1858

<sup>27</sup>Hill S.I., Desobry F., Garnsey E.W., Chong Y.-F., 2010, *The Impact on Energy Consumption of Daylight Saving Clock Changes*, Energy Policy 38, s. 4955-4965.

<sup>28</sup>Ferguson S.A., Preusser D.F., Lund A.K., Zador P.L., Ulmer R.G., 1995, *Daylight Saving Time and Motor Vehicle Crashes: The Reduction in Pedestrian and Vehicle Occupant Fatalities*, American Journal of Public Health, Vol. 85, No. 1, s. 92-95.

<sup>29</sup>Aries M.B.C., Newsham G.R., 2008, *Effect of Daylight Saving ...*, str. 1863

ponieważ w modelach regresji liniowej każdemu kwadransowi przypisana była zmienna zerojedynkowa. Nie pojawiał się w nich problem następowania godziny 3:00 po 1:00. Inaczej jest w modelach szacowanych na szeregach czasowych. Jeśli model szacowany jest iteracyjnie, tak jak na przykład modele przestrzeni stanów, wówczas może zachodzić potrzeba bardziej wyszukanego podejścia. Zilustrowaniu zagadnienia służyć może model autoregresyjny rzędu 96 dla danych kwadransowych. W przypadku zmiany czasu prosty model AR(96) nie będzie poprawnie wyjaśniał zmienności modelowanej zmiennej. Wydaje się, że rozwiązaniem tego problemu mogłoby być potraktowanie czterech kwadransów pominiętej godziny drugiej jako braki obserwacji. Podejście takie wymaga rozszerzenia modelu, które jest możliwe w elastycznych modelach estymowane za pomocą filtru Kalmana.

Rozważmy podobnie jak w poprzedniej części pracy model określony równaniami (1.11)-(1.13), z tą różnicą, że zmienna  $y_t$  zawiera braki obserwacji. Na skutek braku wartości zmiennej  $y_t$  w niektórych okresach czasu niezbędne jest wprowadzenie zmian do modelu przestrzeni stanów i zmodyfikowanie filtru Kalmana. Niech  $y_t^\dagger$  będzie zmienną zdefiniowaną dla wszystkich okresów czasu, w tym także dla pominiętych godzin drugich przy zmianie czasu, tak, że  $y_t^\dagger = y_t$  dla punktów czasu, gdy  $y_t$  jest obserwowana i nieobserwowaną w pozostałych przypadkach. W porównaniu do modelu (1.11)-(1.13) pojawia się dodatkowe równanie

$$y_t^\dagger = \mu_t + \varepsilon_t,$$

dla  $t = 1, \dots, \tilde{T}$ . Dodatkowo, ze względu na brak możliwości obserwowania wartości zmiennej  $y_t$  niezbędna jest modyfikacja filtru Kalmana. Niech  $\tau$  będzie zbiorem punktów w czasie dla których zmienna  $y_t$  jest nieobserwowana, to jest  $\tau = \{t : y_t = NaN\}$ . Dla  $t \in \tau$  równania filtru Kalmana mają następującą postać

$$\alpha_{t|t-1} = F\alpha_{t-1|t-1}, \quad (1.17)$$

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q, \quad (1.18)$$

$$\alpha_{t|t} = \alpha_{t|t-1}, \quad (1.19)$$

$$P_{t|t} = P_{t|t-1}. \quad (1.20)$$

Estymacja komponentu trendu  $\mu$  została wykonana w programie MatLab na podstawie własnej implementacji modelu. Funkcja wiarygodności została zmaksymalizowana za pomocą algorytmu wyżarzania symulowanego, a uzyskana wartość funkcji log-wiarygodności równa była 161310,20. Oszacowana wartość wariancji składnika losowego  $\varepsilon$  równa jest  $\hat{\sigma}_\varepsilon^2 = 0,026798$ , przy odchyleniu standardowym tego oszacowania  $4,0251 \cdot 10^{-9}$ . Bardzo zbliżone wartości oszacowań parametrów potwierdzają wniosek wyciągnięty w podrozdziale 1.4 o istnieniu niestacjonarnego komponentu w danych o zapotrzebowaniu na moce przesyłowe krajowego systemu elektroenergetycznego.

## Rozdział 2

### Analiza sezonowości

Na podstawie wyników z poprzedniego rozdziału nie można odrzucić hipotez o istnieniu niestacjonarności i różnych rodzajów sezonowości w rozważanym szeregu. Wykorzystywane wcześniej testy są testami uniwersalnie stosowanymi w analizie sezonowości i stacjonarności danych. Równie jak powszechne jest ich wykorzystanie w analizach miesięcznych i kwartalnych, tak rzadkie wykorzystanie dla danych o wyższej częstotliwości.

Rozszerzony test Dickeya-Fullera pozwala zbadać istnienie pierwiastka jednostkowego jeśli przy uwzględnieniu opóźnionych różnic rozwiązany zostanie problem autokorelacji reszt wynikający z sezonowości. Rozszerzeniem testu Dickeya-Fullera dla zmiennych sezonowych jest test Dickeya, Haszy i Fullera<sup>1</sup>, w którym wnioskowanie opiera się na sezonowo różnicowanym szeregu. Wadą obydwu testów jest analizowanie deterministycznej sezonowości. Test Hylleberga i innych wydaje się być idealnym narzędziem pozwalającym zbadać występowanie zarówno sezonowości, w wersji stochastycznej i deterministycznej, jak i trendu. Podobnie jednak jak w teście Dickeya, Haszy i Fullera, pojawia się dylemat wyboru opóźnienia przy różnicowaniu. O ile w przypadku danych miesięcznych czy kwartalnych opóźnienie przy różnicowaniu jest bezsporne, to w przypadku danych 15-minutowych wybór ten nie jest jednoznaczny. Pierwszym, naturalnym wyborem kroku różnicowania jest przyjęcie 96 różnic, ale w takim przypadku pomija się sezonowości dzienne i miesięczne. Inny wybór kroku różnicowania musiałby pozwolić uzyskać różnice tygodniowe lub roczne. Ponadto, wybór jednej wartości różnic powoduje nieuwzględnienie dwóch pozostałych. Zastosowanie kilkukrotnego różnicowania, dla każdej sezonowości, doprowadzić może z kolei do problemu nadmiernego różnicowania.

W przeprowadzonym w poprzednim rozdziale teście użyte zostało 96-okresowe różnicowanie. Do przeprowadzenia testu HEGY, który pozwalałby badać 3 rodzaje sezonowości w danych 15-minutowych, istniałaby potrzeba dysponowania komputerem zdolnym przetwarzać i przechowywać w pamięci operacyjnej macierze rozmiaru  $T \times 35040$ . Przy za-

---

<sup>1</sup>Dickey D.A., Hasza D.P., Fuller W. A., 1984, *Testing for Unit Roots in Seasonal Time Series*, Journal of American Statistical Association, 79, s. 355-367.

łożeniu wielkości jednej liczby równej 8 bajtów pamięć operacyjna komputera musiałaby być równa *liczba lat*  $\times$  9,1478 Gb. O ile iteracyjne mnożenie macierzy, choć długotrwałe, mogłoby pozwolić oszacować model postaci

$$\Delta_{35040}y_t = \sum_{i=1}^{35040} a_i Q_{i,t} + \sum_{i=1}^{35040} b_i Y_{i,t-1} + \sum_{i=1}^k c_i \Delta_{35040}y_{t-i} + \varepsilon_t, \quad (2.1)$$

o tyle przeprowadzenie eksperymentu Monte Carlo do wygenerowania wartości krytycznych wydaje się być zajęciem, na które poświęcić trzeba by zbyt wiele czasu. Testy stacjonarności pozwoliły zweryfikować dodatkową hipotezę o istnieniu tendencji wzrostowej w analizowanych danych, jednak nie są wystarczającymi narzędziami do badania złożonej struktury sezonowej. W związku z tym niezbędne było poszukiwanie innych narzędzi.

Dalsze badania nad strukturą wahań sezonowych zostały oparte na modelach szeregów czasowych. W kanonie tych modeli znajduje się model wygładzania wykładniczego, który jest normą w teorii i praktyce. Model wygładzania wykładniczego jest konstrukcją, na której bazują inne bardziej złożone formy prognostyczne. Jak zauważył Taylor<sup>2</sup>, modele wygładzania wykładniczego są popularnym podejściem do prognozowania sezonowych szeregów czasowych. Modele te są szeroko wykorzystywane ze względu na dobre własności prognostyczne, w wielu dziedzinach takich jak kontrolowanie zapasów. Sugeruje to, że metoda Holta-Wintersa może być rozsądnym kandydatem do prognozowania popytu na energię elektryczną. Dodatkowym argumentem może być także wykorzystanie tego typu modeli w innych pracach doktorskich poświęconych tematyce prognozowania popytu na energię elektryczną - na przykład Tomaszewski<sup>3</sup>, Lichota<sup>4</sup> czy Szmit<sup>5</sup>

## 2.1 Model Holta-Wintersa

Holt<sup>6</sup> zaproponował sposób modelowania szeregu czasowego z trendem, który z kolei został rozszerzony przez Wintersa<sup>7</sup> o fluktuacje sezonowe. Model Holta-Wintersa, będący modelem wygładzania wykładniczego dla szeregów zawierających komponenty trendu i sezonowości, jest dobrym standardem w ekonometrii. Model w postaci addytywnej przyj-

<sup>2</sup>Taylor J.W., 2003, *Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing*, Journal of Operational Research Society, 54, s. 799-805.

<sup>3</sup>Tomaszewski M., 2005, *Model przedsiębiorstwa dystrybucyjnego działającego na otwartym rynku energii elektrycznej*, Rozprawa doktorska, Politechnika Opolska, Opole, s. 122.

<sup>4</sup>Lichota A., 2006, *Prognozowanie krótkoterminowe na lokalnym rynku energii elektrycznej*, Rozprawa doktorska, Akademia Górniczo-Hutnicza, Kraków, s. 176.

<sup>5</sup>Szmit A., 2002, *Prognozowanie zapotrzebowania na energię elektryczną. Studium empiryczne dla regionu łódzkiego*, Rozprawa doktorska, Politechnika Łódzka, Łódź.

<sup>6</sup>Holt C.C., 1957, *Forecasting Trends and Seasonal by Exponentially Weighted Averages*, Office of Naval Research Memorandum, 52. przedrukowane w Holt C.C., 2004, *Forecasting Trends and Seasonal by Exponentially Weighted Averages*, International Journal of Forecasting, 20 (1), s. 5-10.

<sup>7</sup>Winters P.R., 1960, *Forecasting Sales by Exponentially Weighted Moving Averages*, Management Science, 6 (3), s. 324-342.



muje postać:

$$\hat{y}_{t+1|t} = l_t + b_t + s_{t-m}, \quad (2.2)$$

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_t + b_t), \quad (2.3)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}, \quad (2.4)$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}, \quad (2.5)$$

gdzie  $l$  jest komponentem poziomym,  $b$  - komponentem trendu,  $s$  - komponentem sezonowym, natomiast  $m$  jest stałą równą okresowi sezonowości. Model szacowany jest za pomocą minimalizacji reszt

$$\min_{\alpha, \beta, \gamma} \left\{ \sum_{i=1}^T (y_t - \hat{y}_{t|t-1})^2 \right\} = \min_{\alpha, \beta, \gamma} \left\{ \sum_{i=1}^T (y_t - (l_{t-1} + b_{t-1} + s_{t-m}))^2 \right\},$$

przy ograniczeniu  $0 < \alpha, \beta, \gamma < 1$ . Przystępny opis modeli wykładniczego znaleźć można w Hydman i Athanasopoulos (2014).

Model postaci (2.2)-(2.5) oszacowany został w środowisku R przy użyciu pakietu forecast. Zakres danych pokrywał dwuletni okres 2011-2012, czyli 70176 kwadranse. Wybór podyktowany był czasem estymacji. Oszacowania parametrów modelu i kryteria informacyjne przedstawione są w tablicy (2.1).

Tablica 2.1: Oszacowania modelu Holta-Wintersa

$\alpha$	$\beta$	$\gamma$	AIC	BIC	loglik
0,92	0,43	0,01	54152,34	54186,21	-27072,17

Źródło: Opracowanie własne.

Reszty w liniowym gaussowskim modelu powinny charakteryzować się trzema właściwościami. Commandeur i Koopman<sup>8</sup> wymieniają je w malejącej skali ważności, tj. niezależność, homoskedastyczność i normalność. Sumaryczny opis testów diagnostycznych znajduje się w aneksie.

Tablica 2.2: Testy diagnostyczne modelu Holta-Wintersa

	statystyka	wartość	p-value	założenie
<i>niezależność</i>	$Q(96)$	24905,550	0	—
	$r(1)$	0,878	0,349	+
	$r(96 * 7)$	24905,550	0	—
<i>homoskedastyczność</i>	$H$	1,034	0,005	—
<i>normalność</i>	$N$	7600,654	0	—

Źródło: Opracowanie własne.

<sup>8</sup>Commandeur J.J.F., Koopman S.J., 2007, *An Introduction to State Space Analysis*, Oxford, str. 14.

Publikowanie diagnostyki reszt nie jest powszechne wśród autorów artykułów, w tym tych poświęconych modelowaniu popytu na energię. Niespełnienie najważniejszego założenia świadczy o występowaniu autokorelacji reszt, czyli nie wytłumaczonej przez model części wahań. Przyczyną może być m.in. występowanie w szeregu dodatkowej wersji sezonowości. Rozwiązaniem problemu autokorelacji reszt może być uwzględnienie dodatkowych sezonowości, do czego potrzebny jest bardziej skomplikowany model.

## 2.2 Model Taylora

Problem występowania dwóch rodzajów fluktuacji czasowych został dostrzeżony w literaturze przez innych autorów. W dziedzinie prognozowania zapotrzebowania na energię elektryczną tę tematykę poruszali Harvey i Koopman<sup>9</sup>, Harvey i inni<sup>10</sup>, Pedregal i Young<sup>11</sup> oraz Dordonnat i inni<sup>12</sup>. Taylor<sup>13</sup> przedstawił model będący rozszerzeniem klasycznego modelu Holta-Wintersa o drugą zmienną sezonową. Na podstawie danych półgodzinnych o popycie na energię elektryczną zauważył wyższość swojego modelu nad standardowym modelem Holta-Wintersa.

Model zaproponowany przez Taylora, czyli model Holta-Wintersa z podwójną sezonowością, przyjmuje postać:

$$\hat{y}_{t+1|t} = l_t + b_t + s_{t-m_1}^{(1)} + s_{t-m_2}^{(2)}, \quad (2.6)$$

$$l_t = \alpha(y_t - s_{t-m_1}^{(1)} - s_{t-m_2}^{(2)}) + (1 - \alpha)(l_t + b_t), \quad (2.7)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}, \quad (2.8)$$

$$s_t^{(1)} = \gamma(y_t - l_{t-1} - b_{t-1} - s_{t-m_2}^{(2)}) + (1 - \gamma)s_{t-m_1}^{(1)}, \quad (2.9)$$

$$s_t^{(2)} = \omega(y_t - l_{t-1} - b_{t-1} - s_{t-m_1}^{(1)}) + (1 - \omega)s_{t-m_2}^{(2)}. \quad (2.10)$$

Pakiet forecast do środowiska R umożliwia oszacowanie także modelu Taylora podwójnej sezonowości. Jego oszacowania przedstawione są w tablicy (2.3).

<sup>9</sup>Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity Demand Using Time-varying Splines*, Journal of the American Statistical Association, 88, 1228-1237.

<sup>10</sup>Harvey A.C., Koopman S.J., Riani M., 1997, *The Modeling and Seasonal Adjustment of Weekly Observations*, Journal of Business & Economic Statistics, Vol. 15(3), s. 354-368.

<sup>11</sup>Pedregal D.C., Young P.C., 2006, *Modulated Cycles, an Approach to Modelling Periodic Components from Rapidly Sampled Data*, International Journal of Forecasting, 22 (1), s. 181-194.

<sup>12</sup>Dordonnat V., Koopman S.J., Ooms M., Dessertaine A., 2008, *An Hourly Periodic State Space Model for Modeling French National Electricity Load*,

<sup>13</sup>Taylor J.W., 2003, *Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing*, Journal of Operational Research Society, 54, s. 799-805.

Tablica 2.3: Oszacowania modelu Taylora

$\alpha$	$\beta$	$\gamma$	$\omega$
0,704	$2,35 \cdot 10^{-9}$	0,848	0,679

Źródło: Opracowanie własne.

Analiza reszt, zaprezentowana w tablicy 2.4, zaprzecza hipotezie o braku autokorelacji reszt. Oprócz występowania istotnej statystycznie autokorelacji pierwszego rzędu, reszty z modelu wykazują korelację ze swoimi opóźnionymi o tydzień wartościami. Rozszerzenie modelu o drugą sezonowość nie rozwiązało problemu autokorelacji 672-rzędu. Niespełnienie założenia o homoskedastyczności wraz z wnioskiem o autokorelacji oznacza, że należy spróbować znaleźć lepszą specyfikację modelu, dla której ten problem nie będzie występował.

Tablica 2.4: Testy diagnostyczne modelu Taylora

	statystyka	wartość	p-value	założenie
<i>niezależność</i>	$Q(96)$	69569,209	0	—
	$r(1)$	12271,640	0	—
	$r(96 * 7)$	69569,209	0	—
<i>homoskedastyczność</i>	$H$	1,138	0	—
<i>normalność</i>	$N$	3582,789	0	—

Źródło: Opracowanie własne.

Jak zauważyli Hyndman i inni<sup>14</sup>, niektóre szeregi czasowe, jak na przykład liczba rozmów przychodzących w bankowości detalicznej mierzona w 5-minutowych przedziałach czasu, może przejawiać dodatkowe roczne wahania sezonowe. Kolejnymi przytoczonymi tam przykładami są: dzienna liczba przyjęć szpitalnych, wypłaty gotówkowe z bankomatów, wykorzystanie energii elektrycznej czy zapotrzebowanie na wodę. Daje to podstawę, wraz z wynikami weryfikacji założeń, do analizowania występowania dodatkowego rodzaju wahań sezonowych.

## 2.3 Model Hyndmana i innych

Taylor<sup>15</sup> przedstawił propozycję uwzględniania trzech rodzajów sezonowości w modelowaniu szeregów czasowych. Jakkolwiek model ten jest łatwo zrozumiały i względnie nieskomplikowany w interpretacji, to wydaje się, że jego słabym punktem jest brak możliwości

<sup>14</sup>Hyndman R.J., De Livera A. M., Snyder R.D., 2010, *Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing*, Monash University, Working Paper 15/09, str. 3.

<sup>15</sup>Taylor J.W., 2010, *Triple Seasonal Methods for Short-term Load Forecasting*, European Journal of Operational Research, 204, s. 139-152.

uwzględniania efektów kalendarzowych i zmian czasu. Model Hyndmana i innych<sup>16</sup> jest bardziej złożony niż model Taylora<sup>17</sup> i jest oparty na modelach przestrzeni stanów. Autorzy ci twierdzą także, że ich model przestrzeni stanów oparty o innowacje (*innovation state space model*) z pojedynczym źródłem błędu charakteryzuje się większą prostotą niż klasyczny model przestrzeni stanów (z wielorakim źródłem błędu).

Model Hyndmana i innych w postaci przestrzeni stanów oparty o innowacje zapisany może być w następującej postaci:

$$\hat{y}_t = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t, \quad (2.11)$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t, \quad (2.12)$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t, \quad (2.13)$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t, \quad (2.14)$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad (2.15)$$

gdzie  $m_1, \dots, m_T$  oznaczają okresy poszczególnych sezonowości,  $l_t$  jest komponentem poziomu w okresie  $t$ ,  $b_t$  - gasnącego trendu<sup>18</sup>, natomiast  $b$  jest długookresowym trendem. Komponent  $s_t^{(i)}$  reprezentuje  $i$ -tą sezonowość,  $d_t$  jest procesem ARMA( $p, q$ ), a  $\varepsilon_t$  jest procesem białego szumu o rozkładzie normalnym, ze średnią 0 i wariancją  $\sigma^2$ . Ze względu na elastyczność i zwartość model Hyndmana wykorzystuje trygonometryczną reprezentację sezonowości, która ma postać:

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}, \quad (2.16)$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos(\lambda_j^{(i)}) + s_{j,t-1}^{*(i)} \sin(\lambda_j^{(i)}) + \gamma_1^{(i)} d_t, \quad (2.17)$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin(\lambda_j^{(i)}) + s_{j,t-1}^{*(i)} \cos(\lambda_j^{(i)}) + \gamma_2^{(i)} d_t, \quad (2.18)$$

gdzie  $\gamma_1^{(i)}$  i  $\gamma_2^{(i)}$  są parametrami wygładzania, a  $\lambda_j^{(i)} = 2\pi j/m_i$ . Liczba harmonik potrzebna do modelowania  $i$ -tej sezonowości jest oznaczana przez  $k_i$ . Komponent  $s^{*(i)}$  pojawia się z powodów konstrukcyjnych a jego interpretacja nie jest szczególnie ważna<sup>19</sup>.

<sup>16</sup>Hyndman R.J., De Livera A. M., Snyder R.D., 2010, *Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing*, Monash University, Working Paper 15/09

<sup>17</sup>Taylor J.W., 2010, *Triple Seasonal Methods for Short-term Load Forecasting*, European Journal of Operational Research, 204, s. 139-152.

<sup>18</sup>Szczegółowe opisy modeli zawierających gasnący trend znaleźć można w Gardner E.S.J., McKenzie E., 1985, *Forecasting Trends in Time Series*, Management Science, 31, s. 1237-1246 lub w Taylor J.W., 2003, *Exponential Smoothing with a Damped Multiplicative Trend*, International Journal of Forecasting, 19, s. 715-725.

<sup>19</sup>Harvey A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, str. 42.

Tablica 2.5: Testy diagnostyczne modelu Hyndmana i innych

	statystyka	wartość	p-value	założenie
<i>niezależność</i>	$Q(96)$	25578,949	0	—
	$r(1)$	52,710	3,867e-013	—
	$r(96 * 7)$	25578,949	0	—
<i>homoskedastyczność</i>	$H$	1,060	4,299e-006	—
<i>normalność</i>	$N$	6844,670	0	—

Źródło: Opracowanie własne.

Model Hyndmana i innych został oszacowany za pomocą polecenia `tbats` z pakietu `forecast` środowiska R. Wyniki testowania hipotez dotyczących założeń modelu przedstawione w tablicy 2.5 zmuszają do odrzucenia hipotezy o braku autokorelacji pierwszego rzędu. Hipoteza o braku autokorelacji aż do 96 opóźnień także jest odrzucana. Odnotować należy korelację bieżących reszt ze swoimi opóźnionymi o tydzień wartościami. Dodatkowo niespełnione są założenia o stałości wariancji reszt i normalności ich rozkładu. Na podstawie analizy własności reszt należy stwierdzić, że niezbędne są dalsze prace nad modelowaniem sezonowości w szeregu zapotrzebowania na moce przesyłowe krajowej sieci elektroenergetycznej. Wydaje się, że w szeregu popytu na energię elektryczną czy zapotrzebowania na moce przesyłowe może występować jednocześnie sezonowość stochastyczna i deterministyczna. Uwzględniając wszystkie wnioski z tego podrozdziału warte rozważenia wydają się być klasyczne modele przestrzeni stanów, które pozwolą modelować zarówno deterministyczne jak i stochastyczne wersje sezonowości.

## 2.4 Modele ze stochastyczną i deterministyczną sezonowością

Analizy wykonane przy użyciu modeli zaprezentowanych w poprzednim rozdziale potwierdzają postawione wcześniej hipotezy o niestacjonarności rozważanego szeregu. Uwzględnienie trzech rodzajów sezonowości jest uargumentowane w podrozdziałach 2.1-2.3. Zgodnie z uzyskanymi wynikami szereg zapotrzebowania na moce przesyłowe zawiera stochastyczny trend, wobec tego zaproponowany został model łączący składnik stochastycznego trendu z trzema różnymi rodzajami sezonowości. Autokorelacji reszt w modelu Hyndmana i innych wskazuje na potrzebę modelowania sezonowości jeszcze w jakiś inny, odmienny od przedstawionych, sposób. Wykorzystanie klasycznego modelu przestrzeni stanów z wielorakim źródłem innowacji jest rozwiązaniem. Skutkiem powyższego może być dłuższy czas estymacji. W niniejszym podrozdziale zaprezentowane zostały wyniki kilku modeli, nierozważanych dotąd w opublikowanych analizach popytu na energię elektryczną w Polsce.

Zaproponowany model jest połączeniem modelu stochastycznego trendu z modelem sto-

chastycznej sezonowości. Obserwowana zmienna jest sumą komponentów trendu, trzech komponentów sezonowych i błędu losowego. Oryginalnym pomysłem na modelowanie trzech rodzajów sezonowości w 15-minutowych danych popytu na energię elektryczną był model postaci:

$$y_t = \mu_t + \gamma_t + \theta_t + \delta_t + \varepsilon_t, \quad (2.19)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (2.20)$$

$$\nu_t = \nu_t + \zeta_t, \quad (2.21)$$

$$\gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \dots - \gamma_{t-95} + \omega_t, \quad (2.22)$$

$$\theta_t = \phi_t(-\theta_{t-1} - \theta_{t-2} - \dots - \theta_{t-6}) + (1 - \phi_t)\theta_t + \eta_t, \quad (2.23)$$

$$\delta_t = \varphi_t(-\delta_{t-1} - \delta_{t-2} - \dots - \delta_{t-11}) + (1 - \varphi_t)\delta_t + \xi_t, \quad (2.24)$$

dla  $i = 1, 2, \dots, T$ , gdzie  $\mu$  jest komponentem poziomym,  $\nu$  - nachylenia, a obydwie modelują stochastyczny trend, natomiast komponenty  $\gamma, \theta, \delta$  reprezentują wahania sezonowe przypisane odpowiednio poszczególnym kwadransom, dniom i miesiącom. Błąd losowy  $\zeta_t$  odpowiada za zmiany nachylenia komponentu trendu. Wzrostowi wariancji  $\sigma_\zeta^2$  odpowiada zwiększenie stochastycznych zmian w trendzie, natomiast w przypadku gdy  $\sigma_\zeta^2 = 0$  trend przyjmuje postać liniowego trendu deterministycznego. Błędy losowe  $\omega, \eta, \xi$ , w równaniach (2.19)-(2.24), są komponentami nieregularnymi, czyli białymi szumami o rozkładzie normalnym z zerowymi średnimi i wariancjami  $\sigma_\omega^2, \sigma_\eta^2, \sigma_\xi^2$ .

Niestety, już w przypadku estymacji modelowania samej tylko sezonowości przypisanej do poszczególnych kwadransów, macierz przejścia w filtrze Kalmana ma rozmiar  $95 \times 95$ . Przy liczbie obserwacji 70176, dla danych za lata 2011-2012, wyznaczenie wartości funkcji wiarygodności dla jednego punktu trwa 4,5 minuty, przez co proces maksymalizacji jest zbyt długotrwały. Z tego powodu, zgodnie z poprzednim podrozdziałem, w części prezentowanych modeli zastosowana została harmoniczna forma sezonowości. Wówczas model (2.19)-(2.24) przyjmuje postać:

$$y_t = \mu_t + \sum_{i=1}^k \gamma_{i,t} + \theta_t + \delta_t + \varepsilon_t, \quad (2.25)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (2.26)$$

$$\nu_t = \nu_t + \zeta_t, \quad (2.27)$$

$$\gamma_{i,t} = \gamma_{i,t-1} \cos(\lambda_i) + \gamma_{i,t-1}^* \sin(\lambda_i) + \omega_{i,t}, \quad (2.28)$$

$$\gamma_{i,t}^* = -\gamma_{i,t-1} \sin(\lambda_i) + \gamma_{i,t-1}^* \cos(\lambda_i) + \omega_{i,t}^*, \quad (2.29)$$

$$\theta_t = \phi_t(-\theta_{t-1} - \theta_{t-2} - \dots - \theta_{t-6}) + (1 - \phi_t)\theta_t + \eta_t, \quad (2.30)$$

$$\delta_t = \varphi_t(-\delta_{t-1} - \delta_{t-2} - \dots - \delta_{t-11}) + (1 - \varphi_t)\delta_t + \xi_t, \quad (2.31)$$

dla  $i = 1, 2, \dots, k$ , gdzie  $k$  jest liczbą harmonik niezbędną do przybliżenia fluktuacji 15-minutowych. Komponent  $\gamma^*$  pełni tę samą rolę co  $s^{*(i)}$  w modelu (2.11)-(2.18). W

praktyce, jak zauważył Harvey<sup>20</sup>, zakłada się, że

$$\text{Var}(\omega_{i,t}) = \text{Var}(\omega_{i,t}^*) = \sigma_i^2 = \sigma_\omega^2,$$

dla  $i = 1, 2, \dots, k$ , co pozwala istotnie zredukować czas estymacji przy niewielkiej stracie w dopasowaniu.

### 2.4.1 Szczegółowa specyfikacja modelu

Zanim zaprezentuję wyniki modeli z różnymi rodzajami sezonowości chciałbym zwrócić uwagę na istotne szczegóły estymacji. W modelu (2.19)-(2.24) niezbędne jest powiązanie parametrów wariancji  $\sigma_\omega^2$  i  $\sigma_\zeta^2$ . Jeśli nie, to może się zdarzyć, że zwiększająca się wariancja  $\omega_t$  może pomniejszać wariancję błędu  $\varepsilon_t$  i tym samym przejmować część wahań błędu z równania obserwacyjnego. Innymi słowy, wariancja przypisywana komponentowi nieregularnemu, czyli błędowi losowemu z równania obserwacyjnego może dążyć do zera,  $\sigma_\varepsilon^2 \rightarrow 0$ . Nie jest to pożądane, ponieważ nie można w takim przypadku określić jaka część wahań ma być przypisywana stochastycznemu komponentowi sezonowości a jaka błędowi losowemu w równaniu pomiarowym. W związku z tym, relacje parametrów  $q_\omega = \sigma_\omega^2/\sigma_\varepsilon^2$  oraz  $q_\zeta = \sigma_\zeta^2/\sigma_\varepsilon^2$  muszą być ustalone na odpowiednim poziomie. Relacje te mogą być wyznaczone na podstawie funkcji wzmocnienia filtra dla komponentów sezonowych i trendu.

Opracowania wykorzystujące modele z trendem i sezonowością nie zawierają wielu przykładów dyskusji nad relacją między błędami losowymi. Problem relacji między wariancjami  $\sigma_\omega^2$  i  $\sigma_\zeta^2$  pojawia się w publikacjach dotyczących filtra Hodricka-Prescotta, tak jak na przykład w Harvey'u i Jaegerze<sup>21</sup> czy Harvey'u i Trimburze<sup>22</sup>. W związku z tym zdecydowałem przedstawić jak ustalenie zależności między tymi parametrami wpływa na własności modelu oraz zaprezentować strategię estymacji modeli z trendem i sezonowością.

Wyznaczenie wpływu relacji między wariancjami błędów losowych rozpocznę od rozwiązania modelu (2.19)-(2.24) ograniczonego do jednego rodzaju wahań sezonowych postaci:

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad (2.32)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (2.33)$$

$$\nu_t = \nu_t + \zeta_t, \quad (2.34)$$

$$\gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \dots - \gamma_{t-s+1} + \omega_{i,t}, \quad (2.35)$$

gdzie  $t = 1, 2, \dots, T$ . Dla uproszczenia wywodu założmy miesięczną częstotliwość obser-

<sup>20</sup>Harvey A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, str. 126 oraz str. 180.

<sup>21</sup>Harvey A.C., Jaeger A., 1993, *Detrending, Stylized Facts and the Business Cycle*, Journal of Applied Econometrics, 8(3), s. 231-247.

<sup>22</sup>Harvey A.C., Trimbur T., 2008, *Trend Estimation and the Hodrick-Prescott Filter*, Journal of the Japan Statistical Society, 38(1), s. 41-49.

wowanej zmiennej, tj.  $s = 12$ . Wówczas, komponent trendu można przedstawić jako:

$$\mu_t = \frac{\zeta_t}{(1-L)^2},$$

a funkcję generującą kowariancję następująco:

$$g_\mu(L) = \frac{1}{(1-L)^2} \frac{1}{(1-L^{-1})^2} \sigma_\zeta^2 = h_\mu(L) \sigma_\zeta^2.$$

Komponent sezonowy może być przedstawiony za pomocą równania:

$$\gamma = \frac{(1-L)}{(1-L^{12})} \omega_t,$$

a jego funkcja generująca kowariancję jako<sup>23</sup>:

$$g_\gamma(L) = \frac{(1-L)}{(1-L^{12})} \frac{(1-L^{-1})}{(1-L^{-12})} \sigma_\omega^2 = h_\gamma(L) \sigma_\omega^2.$$

Funkcję wzmocnienia filtra wyznacza się podstawiając  $L = e^{i\lambda}$ . Funkcja zysku z filtra dla komponentu sezonowego to:

$$\begin{aligned} GL &= \frac{g_\mu(\lambda)}{\sigma_\varepsilon^2 + g_\mu(\lambda) + g_\gamma(\lambda)} = \frac{q_\zeta h_\mu(\lambda)}{1 + q_\zeta h_\mu(\lambda) + q_\omega h_\gamma(\lambda)} \\ &= \frac{q_\zeta \frac{1}{(2-2\cos\lambda)^2}}{1 + q_\zeta \frac{1}{(2-2\cos\lambda)^2} + q_\omega \frac{(2-2\cos 12\lambda)}{(2-2\cos\lambda)}}, \end{aligned} \quad (2.36)$$

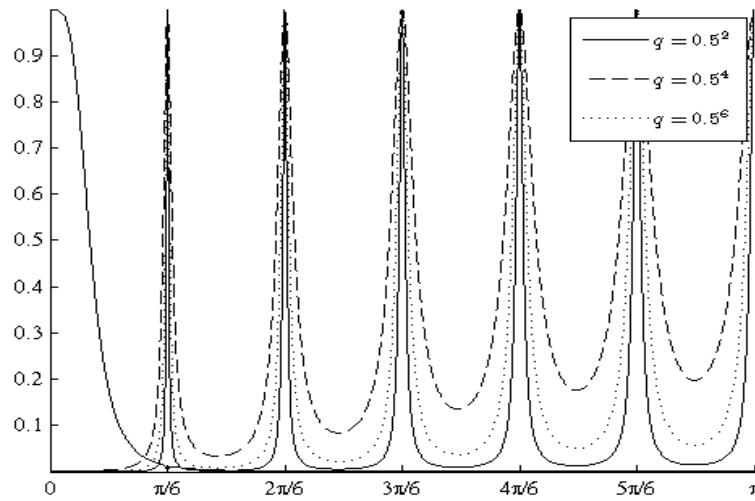
$$\begin{aligned} GS &= \frac{g_\gamma(\lambda)}{\sigma_\varepsilon^2 + g_\mu(\lambda) + g_\gamma(\lambda)} = \frac{q_\omega h_\gamma(\lambda)}{1 + q_\zeta h_\mu(\lambda) + q_\omega h_\gamma(\lambda)} \\ &= \frac{q_\omega \frac{(2-2\cos 12\lambda)}{(2-2\cos\lambda)}}{1 + q_\zeta \frac{1}{(2-2\cos\lambda)^2} + q_\omega \frac{(2-2\cos 12\lambda)}{(2-2\cos\lambda)}}, \end{aligned} \quad (2.37)$$

gdzie  $q_\zeta = \frac{\sigma_\zeta^2}{\sigma_\varepsilon^2}$  oraz  $q_\omega = \frac{\sigma_\omega^2}{\sigma_\varepsilon^2}$ . Funkcje wzmocnienia filtra mają inny przebieg w zależności od parametrów  $q_\zeta$  i  $q_\omega$ . Wykres 2.1 przedstawia funkcje wzmocnienia filtra dla  $q_\zeta$  ustalonego na poziomie  $q_\zeta = 0.001$ , oraz dla różnych wartości  $q_\omega$ .

---

<sup>23</sup>Na podstawie wzoru  $(1+L+L^2+\dots+L^{s-1}) = (1-L)/(1-L^{12})$ .



Rysunek 2.1: Funkcje wzmocnienia dla  $(q_\zeta, q_\omega) \in \{(0,1^3, 0,5^2); (0,1^3, 0,5^4); (0,1^3, 0,5^6)\}$ 

Źródło: Opracowanie własne.

Rozważmy model określony równaniami (2.32)-(2.35) z tą różnicą, że zmienność sezonowa modelowana jest za pomocą funkcji trygonometrycznych zamiast zmiennych zero-jedynkowych:

$$y_t = \mu_t + \sum_{i=1}^k \gamma_{i,t} + \varepsilon_t, \quad (2.38)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (2.39)$$

$$\nu_t = \nu_t + \zeta_t, \quad (2.40)$$

$$\gamma_{i,t} = \gamma_{i,t-1} \cos(\lambda_i) + \gamma_{i,t-1}^* \sin(\lambda_i) + \omega_{i,t}, \quad (2.41)$$

$$\gamma_{i,t}^* = -\gamma_{i,t-1} \sin(\lambda_i) + \gamma_{i,t-1}^* \cos(\lambda_i) + \omega_{i,t}, \quad (2.42)$$

dla  $t = 1, 2, \dots, T$ , gdzie  $\mu$  jest komponentem poziomym,  $\nu$  - nachylenia,  $\gamma$  reprezentuje wahania sezonowe, natomiast  $\varepsilon, \zeta, \omega$  są zaburzeniami losowymi z zerową średnią i wariancjami odpowiednio  $\sigma_\varepsilon^2, \sigma_\zeta^2, \sigma_\omega^2$  i pełnią te same role co w modelu (2.32)-(2.35). Podczas, gdy funkcja generująca autokowariancję dla komponentu długookresowego jest taka sama jak określona równaniem (2.36), natomiast dla komponentu sezonowego ma postać:

$$g_\gamma(\lambda) = \frac{2 - 2 \sin(\lambda_j) \cos(\lambda_j) + 2 \cos(\lambda) [\sin(\lambda_j) - \cos(\lambda_j)]}{4 [\cos(\omega) - \cos(\lambda_j)]^2}. \quad (2.43)$$

Funkcję generującą autokowariancję można łatwo uzyskać z macierzowego zapisu trygonometrycznej sezonowości:

$$\begin{bmatrix} \gamma_t \\ \gamma_t^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{t-1} \\ \gamma_{t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_t \\ \omega_t^* \end{bmatrix}. \quad (2.44)$$

Powyższe równanie może być zapisane przy wykorzystaniu operatora opóźnień w postaci:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_t \\ \gamma_t^* \end{bmatrix} - \begin{bmatrix} L \cos \lambda_j & L \sin \lambda_j \\ -L \sin \lambda_j & L \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_t \\ \gamma_t^* \end{bmatrix} = \begin{bmatrix} \omega_t \\ \omega_t^* \end{bmatrix}, \quad (2.45)$$

co sprowadza się do:

$$\begin{bmatrix} \gamma_t \\ \gamma_t^* \end{bmatrix} = \frac{1}{1 - 2L \cos(\lambda_j) + L^2} \begin{bmatrix} 1 - L \cos \lambda_j & L \sin \lambda_j \\ -L \sin \lambda_j & 1 - L \cos \lambda_j \end{bmatrix} \begin{bmatrix} \omega_t \\ \omega_t^* \end{bmatrix}, \quad (2.46)$$

wobec czego:

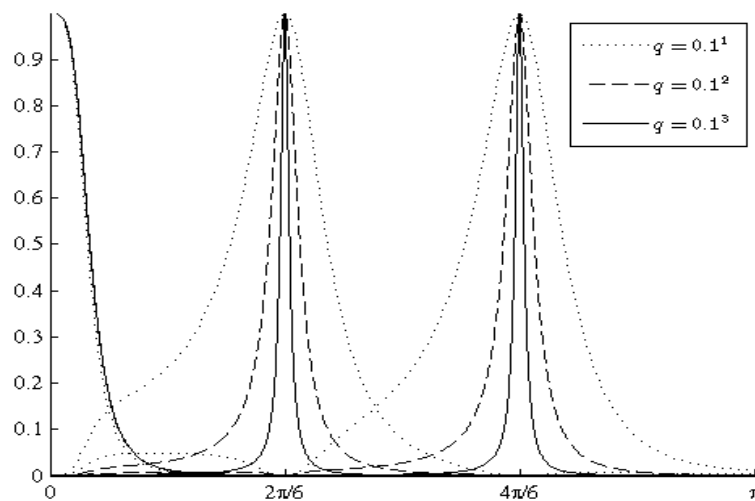
$$\gamma_t = \frac{(1 - L \cos(\lambda_j))\omega_t + L \sin(\lambda_j)\omega_t^*}{1 - 2L \cos(\lambda_j) + L^2}. \quad (2.47)$$

Przy założeniu równej wariancji błędów  $\omega_t$  i  $\omega_t^*$ , analogicznie jak w Harveyu<sup>24</sup>, funkcja generująca autokowariancję ma postać:

$$g_\gamma(L) = \frac{(1 - L \cos(\lambda_j) + L \sin(\lambda_j))(1 - L^{-1} \cos(\lambda_j) + L^{-1} \sin(\lambda_j))}{(1 - 2L \cos(\lambda_j) + L^2)(1 - 2L^{-1} \cos(\lambda_j) + L^{-2})} \sigma_\omega^2. \quad (2.48)$$

Wykres 2.2 przedstawia funkcje wzmocnienia komponentu trendu oraz sezonowości modelowanej przy użyciu dwóch częstości  $\frac{2\pi}{6}$  oraz  $\frac{4\pi}{6}$ . Za pomocą ilorazów wariancji  $q_\zeta$  i  $q_\omega$  można kontrolować stromość funkcji wzmocnienia. Im mniejsze  $q_\omega$ , tym bardziej stroma jest funkcja zysku z filtra dla komponentu sezonowego. Wykres prezentuje trzy przypadki, gdy  $(q_\zeta, q_\omega) = (0, 1^i, 0, 1^i)$  dla  $i = 1, 2, 3$ .

Rysunek 2.2: Funkcje wzmocnienia dla  $(q_\zeta, q_\omega) \in \{(0, 1^1, 0, 1^1); (0, 1^2, 0, 1^2); (0, 1^3, 0, 1^3)\}$



Źródło: Opracowanie własne.

<sup>24</sup>Harvey A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, str. 126 oraz str. 180.

Pokazane zostało jak ustalenie relacji między błędami w równaniu trendu i sezonowości przekłada się na właściwości modelu, więc dalej zaprezentuję koncepcję estymacji modelu klasy (2.32)-(2.35). Strategia ta w literaturze określana jest jako wyprowadzenie (*concentrate out*) jednego z parametrów z funkcji wiarygodności. W pracach, w których autorzy korzystali z filtru Kalmana, pomija się to zagadnienie, a zgodnie z najlepszą wiedzą autora nie istnieje opracowanie w polskiej literaturze poświęcone temu zagadnieniu. Wyprowadzenie parametru z funkcji wiarygodności pozwala uzyskiwać właściwe oszacowania parametrów bez konieczności ustalania relacji wariancji. Problem ten nie pojawiał się przy szacowaniu modelu przestrzeni stanów dla filtru Hodricka-Prescotta, ponieważ w modelu tym relacja wariancji błędów z równań pomiarowego i stanu była ustalona.

Wstępem do opisu zagadnienia wyprowadzania parametru jest przedstawienie logarytmu funkcji wiarygodności modelu przestrzeni stanów, który dany jest wzorem<sup>25</sup>

$$\log L = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log |F_t| - \frac{1}{2} \sum_{t=1}^T v_t' F_t^{-1} v_t, \quad (2.49)$$

gdzie  $v_t = y_t - y_{t|t-1}$ , dla  $t = 1, 2, \dots, T$ , a  $F_t$  jest macierzą kowariancji. Dla przypadku modelu z jedną zmienną funkcja wiarygodności ma postać<sup>26</sup>

$$\log L = -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log f_t - \frac{1}{2} \sum_{t=1}^T v_t' f_t^{-1} v_t. \quad (2.50)$$

Wyprowadzenie wariancji błędu pomiaru polega na ustaleniu pozostałych parametrów wariancji w proporcji do niego. Wtedy  $\text{Var}(\varepsilon_t) = \sigma_*^2 h_t$ , dla np.:  $h_t = 1$ , natomiast macierze postaci przestrzeni stanów modelu i macierze filtru Kalmana muszą zostać przemnożone<sup>27</sup> przez  $\sigma_*^2$ . Harvey<sup>28</sup> zauważył, że taka reparametryzacja jest korzystna, ponieważ wykorzystuje liniowości do zredukowania wymiaru przestrzeni parametrów. Filtr Kalmana może być wówczas zastosowany niezależnie od wyprowadzonej zmiennej. Reparametryzacja modelu uprawnia do zapisania wektora szacowanych parametrów w postaci  $\psi = [\psi_*', \sigma_*^2]'$ , gdzie  $\psi_*$  jest wektorem zawierającym  $n - 1$  parametrów i przedstawienia funkcji (2.50) w postaci

$$\log L = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma_*^2 - \frac{1}{2} \sum_{t=1}^T \log f_t - \frac{1}{2\sigma_*^2} \sum_{t=1}^T \frac{v_t^2}{f_t}. \quad (2.51)$$

Z pochodnej cząstkowej (2.51) otrzymuje się estymator wyprowadzonego parametru  $\sigma_*^2$ .

$$\frac{\partial \log L}{\partial \sigma_*^2} = 0 \Leftrightarrow \tilde{\sigma}_*^2(\psi_*) = \frac{1}{T} \sum_{t=1}^T \frac{v_t^2}{f_t} \quad (2.52)$$

<sup>25</sup>Stosuję oznaczenia przyjęte w opisie modeli przestrzeni stanów i filtru Kalmana w aneksie.

<sup>26</sup>Zastosowanie małych liter podkreślać ma przypadek jednej zmiennej.

<sup>27</sup>Przy zastosowaniu konwencji przyjętej w opisie modeli przestrzeni stanów i filtru Kalmana zapisać należy  $\text{Var}(\zeta) = \sigma_*^2 Q$ ,  $P_{t+1|t} = \sigma_*^2 P_{t+1|t}^*$  i  $f_t = \sigma_*^2 f_t^*$

<sup>28</sup>Harvey A.C., 1989, *Forecasting, Structural Time Series Models ...*, str. 126.

Funkcję wiarygodności z wyprowadzonym parametrem uzyskuje się z (2.51) przy podstawieniu (2.52)

$$\log L = -\frac{T}{2}(\log 2\pi + 1) - \frac{1}{2} \sum_{t=1}^T \log f_t - \frac{T}{2} \log \frac{1}{T} \sum_{t=1}^T \frac{v_t^2}{f_t}. \quad (2.53)$$

Zaprezentowane dalej modele zostały oszacowane przy wyprowadzeniu parametru wariancji błędu losowego z równania pomiarowego  $\sigma_\varepsilon^2$  z funkcją wiarygodności daną równaniem (2.53).

### 2.4.2 Modele z zero-jedynkową sezonowością

Analogicznie jak w podrozdziałach 2.1-2.3, modelowanie danych 15-minutowych rozpoczęte zostało od najprostszego modelu, który był stopniowo rozbudowywany. Przedstawione w tym podrozdziale modele ekonometryczne zostały oszacowane przy wykorzystaniu wyprowadzania parametru z funkcji wiarygodności. Zmienną obserwowaną w tych modelach jest logarytm 15-minutowego zapotrzebowania na moce przesyłowe KSE.

Wyjściowy model stochastycznego trendu z sezonowością przyjmuje formę

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad (2.54)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (2.55)$$

$$\nu_t = \nu_t + \zeta_t, \quad (2.56)$$

$$\gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \dots - \gamma_{t-95} + \omega_t, \quad (2.57)$$

dla  $t = 1, 2, \dots, T$ , gdzie  $\mu$  jest komponentem poziomym,  $\nu$  - nachylenia,  $\gamma$  reprezentuje dzienne wahania sezonowe, natomiast  $\varepsilon$  odpowiada błędowi losowemu. Komponenty  $\zeta, \omega$  są zaburzeniami losowymi z zerową średnią i wariancjami odpowiednio  $\sigma_\zeta^2$  i  $\sigma_\omega^2$ . Został on oszacowany w wersji ze stochastyczną sezonowością, ponieważ wyniki analiz modeli z podrozdziałów 2.1-2.3 przesądziły o tym, że deterministyczna wersja sezonowości nie jest właściwa. Oszacowania modelu prezentuje tablica 2.6 w kolumnie (1). Model ten został porównany ze swoimi rozszerzonymi wersjami za pomocą współczynnika determinacji  $R^2$ , który, zgodnie z Harvey'em<sup>29</sup>, ma postać

$$R^2 = 1 - SSE / \sum_{t=2}^T (\Delta y_t - \overline{\Delta y})^2,$$

gdzie  $SSE = \bar{f} \sum_{t=1}^T \tilde{v}_t^2 = (T-d)\tilde{\sigma}^2$ ,  $d$  jest liczbą komponentów w modelu,  $\overline{\Delta y}$  jest średnią z przyrostów zmiennej  $y$ , natomiast  $\bar{f}$  jest rozwiązaniem stanu ustalonego  $\bar{f} = \lim_{t \rightarrow \infty} f_t$ . Współczynnik determinacji oraz wartość wariancji błędu pomiarowego wskazują na prze-

<sup>29</sup>Harvey A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter...*, str. 268.

Tablica 2.6: Wyniki dla modeli różnych sezonowości zero-jedynkowych

	(1)	(2)	(3)
	Model ze stochastyczną sezonowością 15-minutową	Model ze stochastycznymi sezonowościami: 15-minutową i dzienną	Model ze stochastycznymi sezonowościami: 15-minutową, dzienną i miesięczną
$10^5 \sigma_\varepsilon^2$	1,9994	1,9932	1,9930
$10^5 \sigma_\zeta^2$	1,3843	1,3711	1,3787
$10^5 \sigma_\omega^2$	0,1326	0,1361	0,1345
$10^5 \sigma_\eta^2$		0,0147	0,0005
$10^5 \sigma_\delta^2$			2,2278
$10^{-3} \log L$	2,2307	2,2305	2,2304
$R^2$	0,6512	0,6521	0,6522

Źródło: Opracowanie własne.

wagę modelu ze stochastyczną sezonowością.

Podobnie jak model Taylora<sup>30</sup> jest rozszerzeniem modelu Holta-Wintersa, tak model postaci

$$y_t = \mu_t + \gamma_t + \theta_t + \varepsilon_t, \quad (2.58)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (2.59)$$

$$\nu_t = \nu_t + \zeta_t, \quad (2.60)$$

$$\gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \dots - \gamma_{t-95} + \omega_t, \quad (2.61)$$

$$\theta_t = \phi_t (-\theta_{t-1} - \theta_{t-2} - \dots - \theta_{t-6}) + (1 - \phi_t)\theta_t + \eta_t, \quad (2.62)$$

jest rozszerzeniem (2.54)-(2.57). Równanie (2.58) zawiera drugą zmienną sezonową  $\theta_t$ , która pozwala modelować wahania charakterystyczne dla poszczególnych dni. Struktura autoregresyjna  $\theta_t$  pozwala przypisywać wartości poszczególnym dniom w modelu wysokoczęstotliwościowym. Pomysł na modelowanie drugiej sezonowości w ten sposób nie został wykorzystany dotąd w literaturze. Zaletą tego podejścia jest brak konieczności sztywnego odwoływania się do konkretnych opóźnień, dzięki czemu nie występuje problem zmian czasu omawiany na początku rozdziału 2. Oszacowania parametrów tego modelu znajdują się w tablicy 2.6 w kolumnie (2). Zastosowana została stochastyczna wersja dziennej sezonowości, a wybór ten został potwierdzony przez niezerowe oszacowanie parametru  $\sigma_\eta^2$ .

<sup>30</sup>Taylor J.W., 2003, *Short-term Electricity Demand Forecasting ...*

Współczynnik determinacji ma wartość wyższą niż dla modelu (1), co jest przesłanką do wniosku o tym, że model z podwójną strukturą sezonową jest lepszy do modelowania rozważanych danych 15-minutowych. Nieintuicyjne wartości funkcji wiarygodności uzyskane dla modeli ogólniejszych uniemożliwiają formalne przetestowanie hipotezy o

Model postaci (2.58)-(2.62) pozwalał na modelowanie zapotrzebowania na moce przesyłowe przy uwzględnieniu wahań sezonowych przypisywanych poszczególnym okresom 15-minutowym i poszczególnym dniom tygodnia. Na podstawie graficznej analizy szeregu zauważyć można okresy zmniejszonego, latem, i zwiększonego, zimą, zapotrzebowania na energię elektryczną. Wobec tego zasadnym wydaje się włączenie do modelu (2.58)-(2.62) dodatkowego komponentu  $\delta$  modelującego sezonowość roczną, a rozszerzony w ten sposób model ma postać

$$y_t = \mu_t + \gamma_t + \theta_t + \delta_t + \varepsilon_t, \quad (2.63)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (2.64)$$

$$\nu_t = \nu_t + \zeta_t, \quad (2.65)$$

$$\gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \dots - \gamma_{t-95} + \omega_t, \quad (2.66)$$

$$\theta_t = \phi_t(-\theta_{t-1} - \theta_{t-2} - \dots - \theta_{t-6}) + (1 - \phi_t)\theta_t + \eta_t, \quad (2.67)$$

$$\delta_t = \varphi_t(-\delta_{t-1} - \delta_{t-2} - \dots - \delta_{t-11}) + (1 - \varphi_t)\delta_t + \xi_t, \quad (2.68)$$

dla  $t = 1, 2, \dots, T$ , i gdzie  $\varphi$  jest zmienną zerojedynkową, przyjmującą wartość 1 tylko dla pierwszego kwadransu w miesiącu. Niewątpliwą zaletą modelu (2.63)-(2.68) jest uwzględnianie 3 rodzajów wahań sezonowych. Wadą wydaje się być czas estymacji, ponieważ w każdej iteracji odwracane są macierze  $116 \times 116$ . Współczynnik  $R^2$  jest dla tego modelu najwyższy, jednak zmiana w stosunku do modelu (2.58)-(2.62) jest mała i wydaje się być niezadowalająca<sup>31</sup>. Niedoskonałości te skłaniają do podjęcia próby modelowania wahań sezonowych w inny sposób, który będzie także pozwalał modelować potrójną strukturę sezonową, ale którego czas estymacji będzie krótszy.

### 2.4.3 Modele z trygonometryczną sezonowością

Wykorzystałem trygonometryczną formę sezonowości ze względu na zbyt długi czas estymacji modeli z sezonowościami modelowanymi zmiennymi zero-jedynkowymi. Modele

<sup>31</sup>Nieintuicyjne wartości funkcji wiarygodności, mniejsze dla ogólniejszych modeli, uniemożliwiają formalne porównanie modeli. Szczegółowe analizy procesu maksymalizacji nie dały podstaw do stwierdzenia, że algorytm maksymalizujący nie osiągnął zbieżności. Mniejsze wartości funkcji wiarygodności związane są ze zdefiniowaniem równań modelu w oparciu o zmieniające się w czasie parametry  $\phi$  i  $\varphi$ .

(2.58)-(2.62) oraz (2.63)-(2.68) zapisane mogą być w postaci:

$$y_t = \mu_t + \sum_{i=1}^k \gamma_{i,t} + \varepsilon_t, \quad (2.69)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (2.70)$$

$$\nu_t = \nu_t + \zeta_t, \quad (2.71)$$

$$\gamma_{i,t} = \gamma_{i,t-1} \cos(\lambda_i) + \gamma_{i,t-1}^* \sin(\lambda_i) + \omega_{i,t}, \quad (2.72)$$

$$\gamma_{i,t}^* = -\gamma_{i,t-1} \sin(\lambda_i) + \gamma_{i,t-1}^* \cos(\lambda_i) + \omega_{i,t}^*, \quad (2.73)$$

dla  $i = 1, 2, \dots, k$ , gdzie  $k$  jest liczbą harmonik niezbędną do przybliżenia fluktuacji sezonowych. W praktyce, jak zauważył Harvey<sup>32</sup>, zakłada się, że

$$\text{Var}(\omega_{i,t}) = \text{Var}(\omega_{i,t}^*) = \sigma_{\omega^*}^2 = \sigma_{\omega}^2,$$

dla  $i = 1, 2, \dots, k$ , co pozwala istotnie zredukować czas estymacji przy niewielkiej stracie w dopasowaniu. Im większa liczba harmonik, tym przybliżenie sezonowości jest dokładniejsze. Zarówno prosta postać jak i złożona struktura wahań sezonowych mogą być modelowane w ten sposób.

Dodatkowym, co najmniej równie ważnym, powodem zastosowania funkcji trygonometrycznych jest ich większa elastyczność w modelowaniu sezonowości. Jakkolwiek wydaje się, że w literaturze przedmiotu nie istnieje jednoznaczne wskazanie na wyższość wykorzystania funkcji trygonometrycznych. Dla przykładu Harvey, Koopman i Riani<sup>33</sup> modelując dane tygodniowe użyli obydwu metod uwzględnienia wahań kalendarzowych.

Wybór tylko części harmonik do modelowania sezonowości uzasadniony jest tak jak w Hyndmanie i innych<sup>34</sup>, czasem estymacji. Analiza danych oraz analogiczna procedura do przedstawionej w Hyndmanie i innych<sup>35</sup> pozwoliła ustalić wielkości  $k_d = 7$ ,  $k_t = 6$  i  $k_r = 3$ , gdzie  $k_d, k_t, k_r$  są odpowiednio harmonikami dla wahań dziennych, tygodniowych i rocznych.

Tablica 2.7 prezentuje wyniki estymacji trzech modeli. Wartość współczynnika determinacji na poziomie 82% dla modelu z potrójną strukturą sezonową w porównaniu do 53% dla modelu uwzględniającego tylko sezonowości dzienne i tygodniowe wskazuje na to, że bardziej złożony jest właściwy. Statystyka testowa testu ilorazu wiarygodności hipotezy o nieistotności wahań rocznych wynosi 18410 i znacznie przewyższa wartość krytyczną  $\chi_6^2(0,95) = 12,59$ . Pokazany w rozdziale 1 stochastyczny charakter trendu w analizowanych

<sup>32</sup>Harvey A.C., 1989, *Forecasting, Structural Time Series Models ...*, str. 43.

<sup>33</sup>Harvey A.C., Koopman S.J., Riani M., 1997, *The Modeling and Seasonal Adjustment ...*

<sup>34</sup>Hyndman R.J., De Livera A. M., Snyder R.D., 2010, *Forecasting Time Series with Complex ...*

<sup>35</sup>Hyndman R.J., De Livera A. M., Snyder R.D., 2010, *Forecasting Time Series with Complex ...*, str. 21. Z wybranego fragmentu danych usuwa się liniowy trend, który wykorzystuje się w iteracyjnej procedurze. W każdym kroku procedury wykonuje się regresję danych na coraz to bardziej rozszerzonym zbiorze harmonik. Kolejność dodawania harmonik zależy od ich parametrów z modelu regresji  $y_t = \sum_{i=1}^T \sum_{j=1}^{k_i} a_j^{(i)} \cos(\lambda_j^{(i)} t) + b_j^{(i)} \sin(\lambda_j^{(i)} t)$ .

Tablica 2.7: Wyniki dla modeli różnych sezonowości trygonometrycznych

	(1)	(2)	(3)
	Model z dwoma rodzajami stochastycznymi sezonowościami z trendem stochastycznym	Model z trzema rodzajami stochastycznymi sezonowościami z trendem stochastycznym	Model z trzema rodzajami stochastycznymi sezonowościami z trendem deterministycznym
$10^5 \sigma_\epsilon^2$	2,6916	0,9692	1,3343
$10^5 \sigma_\zeta^2$	3,6358	0,0087	0*
$10^5 \sigma_\omega^2$	3,4673	122,8860	139,1378
$10^{-3} \log L$	2,1351	2,2258	2,2183
$R^2$	0,5314	0,8210	0,7926

\* oznacza, że parametr nie był szacowany, tylko zostało na niego nałożone ograniczenie.

*Źródło:* Opracowanie własne.

danych został potwierdzony wynikami modelu z trzema rodzajami sezonowości z deterministycznym trendem. Testowana testem ilorazu wiarygodności hipoteza  $H_0 : \sigma_\zeta^2 = 0$  także musi zostać odrzucona na poziomie istotności 5%<sup>36</sup>. Stwierdzić zatem należy, że model ze potrójną stochastyczną strukturą sezonową i stochastycznym trendem jest właściwy do modelowania analizowanego szeregu.

Przedstawione powyżej analizy potwierdzają założoną dodatkową hipotezę o istnieniu potrójnej struktury sezonowej w szeregu zapotrzebowania na moce przesyłowe krajowego systemu elektroenergetycznego. Dodatkowo pokazałem, że wahania sezonowe dzienne, tygodniowe i roczne powinny być modelowane w sposób stochastyczny tak jak komponent trendu. Rozdziały 1 i 2 zawierają analizy składowych długo- trendu i krótkookresowych, do których zalicza się sezonowość, natomiast pozostała część wahań, czyli komponent średniookresowy jest przedmiotem rozdziału 3.

<sup>36</sup>Wartość statystyki testowej była równa 1500, natomiast wartość krytyczna  $\chi_1^2 = 3,84$

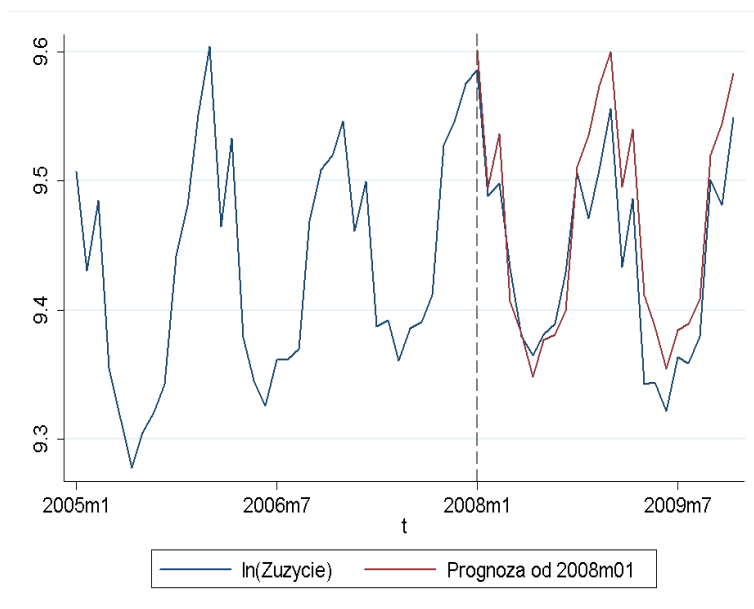


## Rozdział 3

# Wpływ cyklu koniunkturalnego

Zgodnie z teorią analizy szeregów czasowych, każdy z nich możemy zdekomponować na składniki długo-, średnio- oraz krótko-okresowe. Jak już zostało to zaprezentowane w pierwszym rozdziale, w szeregu zapotrzebowania na moce przesyłowe krajowej sieci elektroenergetycznej występuje składowa trendu. Analizie komponentu sezonowego, zaliczanego do wahań krótkookresowych, poświęcony został natomiast rozdział drugi. Składnik średniookresowy utożsamiany jest z tą częścią wahań, która związana jest z cyklem koniunkturalnym, a właściwe jego uwzględnienie i modelowanie może być przydatne również w wyjaśnianiu zmienności szeregu wysokoczęstotliwościowego i w prognozowaniu jego wartości w średnim horyzoncie czasowym.

Rysunek 3.1: Logarytm krajowego zużycia energii elektrycznej i jego prognoza



Źródło: Opracowanie własne.

Rozważmy miesięczny szereg czasowy krajowego zużycia energii elektrycznej dla okresu od stycznia 1993 do sierpnia 2013 roku. Dla próbki obserwacji 1993m01-2007m12, osza-

cowany został model SARIMA<sup>1</sup> dla logarytmu zmiennej, a następnie na jego podstawie wyznaczono prognozę na następne 24 miesiące. Rysunek 3.1 przedstawia wyznaczone prognozy w porównaniu z rzeczywiście zaobserwowanymi wartościami. Prognozowana wartość zużycia znacznie przewyższa zaobserwowaną. Widocznie mniejsze niż przewidywane zużycie może być związane z kryzysem zapoczątkowanym przez problemy na amerykańskim rynku kredytów hipotecznych. Wydaje się, że wiedza o stanie gospodarki, o okresach prosperity i recesji, mogłaby istotnie poprawić prognozy popytu na energię elektryczną.

W związku z tym, poniżej zaprezentowana została weryfikacja hipotezy, że przy wykorzystaniu informacji o wahanich średniookresowych stanu gospodarki uzyskuje się lepsze prognozy popytu na energię elektryczną. Wyliczyłem nie tylko dwie zmienne ilustrujące stan koniunktury, ale także przedstawiłem propozycję szeregu, za pomocą którego prognozować można stan gospodarki w przyszłości. Pokazałem w dalszej części rozdziału, że różne szeregi popytu na energię elektryczną zawierają składowe średniookresowe identyfikowane z cyklem koniunkturalny. Przedstawione poniżej próby i wyniki skwantyfikowania polskiego cyklu koniunkturalnego zostały dalej wykorzystane do zbadania istnienia zależności przyczynowo-skutkowej z szeregami opisującymi zapotrzebowanie na energię elektryczną oraz jej ewentualnego kierunku. Ostatnia część rozdziału poświęcona jest modelowaniu i prognozowaniu danych wysokoczęstotliwościowych z wykorzystaniem wahań cyklicznych jako zmiennej egzogenicznej.

### 3.1 Równoczesny i wyprzedzający wskaźnik koniunktury

Najpowszechniej przytaczana definicja wahań koniunkturalnych pochodzi od Mitchella<sup>2</sup>. Cykl koniunkturalny, zgodnie z nią, jest częścią wahań zagregowanej aktywności gospodarczej w kraju, w którym praca zorganizowana jest w przedsiębiorstwach, przy czym uwzględnia się tylko te rodzaje aktywności, które mają systematyczny i ekonomiczny charakter. Cykl składa się z okresu ekspansji przejawiającego się w tym samym czasie w wielu typach działalności gospodarczej, okresu dekonunktury, recesji i ożywienia łączącego się z okresem ekspansji kolejnego cyklu. Sekwencja tych zmian ma charakter powtarzalny. Czas trwania cyklu waha się od ponad roku do dziesięciu, dwunastu lat. Poszczególne cykle nie mogą być dzielone na krótsze o podobnym charakterze. Według definicji Gaudreault i innych<sup>3</sup> termin „cykl koniunkturalny” odnosi się do wspólnych ruchów w bardzo

<sup>1</sup>Liczby opóźnień w modelu zostały wybrane za pomocą procedury od ogółu do szczegółu i miały postać odpowiednio SARIMA  $(3,0,0) \times (0,1,1)_{12}$ . Model został oszacowany w programie Stata.

<sup>2</sup>Mitchell W.C., 1927, *Business Cycles: The Problem and Its Setting*, National Bureau of Economic Research, Nowy Jork, s. 481.

<sup>3</sup>Gaudreault C., Lamy R., Liu Y., 2003, *New Coincident, Leading and Recession Indexes for the Canadian Economy: An Application of the Stock and Watson Methodology*, Department of Finance, Working Paper 2003-12, str. 2

wielu zmiennych ekonomicznych, takich jak produkcja, zatrudnienie czy handel detaliczny. Zgodnie z powyższym, cykl koniunkturalny jest wspólnym nieobserwowalnym komponentem, który jest zawarty w wielu obserwowanych szeregach ekonomicznych. Przedstawione w nieniejszej pracy modele pozwalają uzyskać wskaźniki zgodne z tym podejściem.

GUS, IBnGR oraz NBP publikują wyniki swoich badań ankietowych dotyczących koniunktury w wybranych sektorach gospodarki. Wskaźniki te prezentują odczucia, oceny oraz przewidywania podmiotów dotyczące gospodarki i koniunktury. Indeks koniunktury dla całej polskiej gospodarki publikowany jest przez OECD, z tym, że jest to wskaźnik wyprzedzający. Biuro BIEC szacuje wskaźniki równoczesne i wyprzedzające, ale nie są one dostępne publicznie. Do listy publikowanych zmiennych doliczyć można eksperymentalne wskaźniki koniunktury, równoczesny i wyprzedzający, GUS<sup>4</sup>. Brak dostępu do publikowanych równoczesnych wskaźników koniunktury był powodem samodzielnego wyliczenia indeksów koniunktury. Do konstrukcji wskaźnika wyprzedzającego OECD wykorzystywane są zmienne, których wartości znane są z opóźnieniem przez co część efektu wyprzedzania jest unieważniana.

Do estymacji własnych wskaźników koniunktury wybrane zostały szeregi opisujące różne części gospodarki. Indeks produkcji przemysłowej odzwierciedlać ma produkcję, szereg wolumenu handlu detalicznego, koniunkturę w handlu, eksport i import, koniunkturę na dobra krajowe i zagraniczne, w końcu bieżący szereg ufności konsumenckiej opisuje sytuację gospodarstw domowych. Szeregami wyprzedzającymi są przewidywana produkcja sprzedana w przetwórstwie przemysłowym w ciągu najbliższych 3 miesięcy, indeks WIG20 oraz spread między rentownością 5-letnich obligacji skarbu państwa a 3 miesięcznym WIBORem. Dokładne nazwy i źródła danych przedstawia tablica 3.1.

Tablica 3.1: Zmienne użyte do estymacji wskaźników koniunktury

Szereg	Początek	Koniec	Źródło	Wskaźnik
Produkcja przemysłowa (2010 = 100)	1991.01	2014.03	OECD	Równoległy
Handel detaliczny (2010 = 100)	1991.01	2014.03	OECD	Równoległy
Import (mld PLN)	2000.01	2014.03	GUS	Równoległy
Export (mld PLN)	2000.01	2014.03	GUS	Równoległy
Bieżący wsk. ufności konsumenckiej	1997.04	2014.03	GUS	Równoległy
Spread między rentownością 5-letnich obligacji skarb. a 3-mies. WIBORem	1999.03	2014.03	NBP	Wyprzedzający
WIG20	1999.03	2014.03	NBP	Wyprzedzający
Oczekiwana wielkość prod. sprzedanej w najbl. 3 mies. w przetwórstwie	1999.03	2010.06	GUS	Wyprzedzający

*Źródło:* Opracowanie własne.

<sup>4</sup>Na dzień 30.06.2014 dane te nie są dostępne publicznie. Moja prośba o przesłanie danych wyłącznie do celów pracy doktorskiej została odrzucona.

Szeregi importu, eksportu oraz przewidywanej produkcji sprzedanej zostały odsezonowane za pomocą metody TRAMO/SEATS dostępnej w programie Eviews. Indeksy produkcji przemysłowej oraz handlu detalicznego dostępne były w wersjach z usuniętymi wahaniami sezonowymi. Bieżący wskaźnik ufności konsumenckiej ma częstotliwość miesięczną tylko od stycznia 2004 roku, wcześniej natomiast publikowany był kwartalnie. Łamana częstotliwość tej zmiennej oraz nierówne zakończenie szeregów nie stanowi przeszkody w konstrukcji wskaźników, gdyż wartości te traktowane są jako braki obserwacji, analogicznie jak w rozdziale pierwszym. Alternatywą, mogłoby być rozszacowanie kwartalnych obserwacji od kwietnia 1997 roku do października 2003 na miesiące jednym ze sposobów rozszacowania zaprezentowanych w Woźniak<sup>5</sup>.

Rozważmy strukturalny model szeregu czasowego, w którym obserwowalna zmienna  $y_t$  jest sumą nieobserwowalnego komponentu trendu  $\mu_{m,t}$ , nieobserwowalnego komponentu cyklicznego  $\psi_{n,t}$  oraz komponentu nieregularnego  $\varepsilon_t$ , czyli białego szumu. Zapisane to może zostać w postaci

$$y_t = \mu_{m,t} + \psi_{n,t} + \varepsilon_t, \quad (3.1)$$

dla  $t = 1, 2, \dots, T$ . Nieobserwowalny komponent  $\mu_{m,t}$  jest zdefiniowany jako trend stochastyczny rzędu  $m$ , dla dodatnich wartości  $m$ , tak, że

$$\mu_{1,t} = \mu_{1,t-1} + \zeta_t, \quad (3.2)$$

$$\mu_{i,t} = \mu_{i,t-1} + \mu_{i-1,t}, \quad (3.3)$$

dla  $i = 2, \dots, m$ , gdzie  $\zeta$  jest szeregiem białego szumu ze średnią zero i stałą wariancją  $\sigma_\zeta^2$ . Trend stochastyczny pierwszego rzędu jest procesem wahań losowych, podczas gdy trend rzędu drugiego jest zintegrowanym procesem błędzenia przypadkowego. Modele trendu mogą przyjąć postać deterministyczną, w przypadku, gdy wariancja zostanie wyestymowana na zerowym poziomie. Wówczas model odpowiada zwykłej regresji tylko ze stałą, lub dwiema zmiennymi – stałą i trendem deterministycznym. Więcej informacji na temat modeli tego typu można znaleźć w Harvey<sup>6</sup> oraz Commandeur i Koopman<sup>7</sup>.

Wykorzystanie nieobserwowalnego komponentu cyklu stochastycznego zostało opisane i rozwinięte przez Harveya<sup>8</sup>, Harveya i Jaegera<sup>9</sup> oraz Trimbura<sup>10</sup>. Nieobserwowalny kom-

<sup>5</sup>Woźniak R., 2011, *Metody rozszacowania kwartalnych indeksów koniunktury konsumenckiej*, Wiadomości Statystyczne, 4, s. 1-10.

<sup>6</sup>Harvey A. C., 1989, *Forecasting, Structural Time Series...*

<sup>7</sup>Commandeur J.F, Koopman S.J., 2007, *An Introduction to State...*

<sup>8</sup>Harvey A. C., 1989, *Forecasting, Structural Time Series...*

<sup>9</sup>Harvey A.C., Jaeger A., 1993, *Detrending, Stylized Facts...*

<sup>10</sup>Trimbur T.M., 2005, *Properties of Higher Order Stochastic Cycles*, Journal of Time Series Analysis, 27(1), s. 1-17.

ponent  $\psi_{n,t}$  jest cyklem stochastycznym rzędu  $n$ , dla dodatnich  $n$ , jeśli

$$\begin{bmatrix} \psi_{i,t} \\ \psi_{i,t}^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{i,t-1} \\ \psi_{i,t-1}^* \end{bmatrix} + \begin{bmatrix} \psi_{i-1,t-1} \\ \psi_{i-1,t-1}^* \end{bmatrix}, \quad (3.4)$$

$$\begin{bmatrix} \psi_{1,t} \\ \psi_{1,t}^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{1,t-1} \\ \psi_{1,t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad (3.5)$$

gdzie  $i = 2, \dots, n$ ,  $\kappa_t \sim N(0, \sigma_\kappa^2)$ ,  $\kappa_t^* \sim N(0, \sigma_{\kappa^*}^2)$ . Parametr  $\rho$  jest parametrem tłumienia (*damping parameter*) i spełnia  $0 < \rho \leq 1$ .

Model wykorzystany do estymacji wskaźników koniunktury jest wielorównaniowym modelem strukturalnego szeregu czasowego ze wspólnym stochastycznym cyklem. Reprezentacja modelu w postaci przestrzeni stanów składa się z dwóch równań, pierwszego, obserwacyjnego dla  $y_t$ , drugiego dla zmiennych stanu opisującego ich dynamikę. Wektor zmiennych stanu  $\alpha_t = (\mu_t', \psi_t')'$ , gdzie  $\mu_t' = (\mu_{m,t}', \mu_{m-1,t}', \dots, \mu_{1,t}')$  i  $\psi_t' = (\psi_{n,t}', \psi_{n-1,t}', \dots, \psi_{1,t}')$ , gdzie  $\mu_{j,t}' = (\mu_{j,t}^{1'}, \mu_{j,t}^{2'}, \dots, \mu_{j,t}^{N'})$ , gdzie superskrypt oznacza komponent trendu  $j$ -tego rzędu dla  $n$ -tej obserwowalnej zmiennej. Równanie pomiarowe ma postać

$$y_t = z_t' \alpha_t + \epsilon_t,$$

dla  $t = 1, \dots, T$ , gdzie pierwszymi  $(m+1)$  elementami wektora  $z_t$  są jedynki, a zera pozostałymi. Równanie przejścia (stanu) dla komponentów trendu jest postaci

$$\mu_t = U_m \mu_{t-1} + i_m \zeta_t,$$

gdzie  $U_m$  jest  $m \times m$  wymiarową, górną trójkątną macierzą jedynek, natomiast  $i_m$ , jest  $m \times 1$  wektorem jedynek. Macierz kowariancji wektora zaburzeń jest równa  $Var(i_t \zeta_t) = i_m i_m' \sigma_\zeta^2$ . Równanie przejścia dla komponentów cyklicznych jest następujące

$$\psi_t = T_\psi \psi_{t-1} + i_n \otimes \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix},$$

gdzie macierz przejścia  $T_\psi$  może być zapisana w postaci

$$T_\psi = I_n \otimes \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} + S_n \otimes I_2,$$

gdzie  $S_n$  jest  $n \times n$  wymiarowa macierzą zer z wyjątkiem elementów diagonalą równym jedynkom, a  $I_2$  jest macierzą identycznościową wymiaru  $2 \times 2$ , natomiast  $\otimes$  oznacza iloczyn Kroneckera. Bardziej szczegółowy opis modelu i jego zapisu w postaci przestrzeni stanów znajduje się w Harvey i Jaeger<sup>11</sup>, Harvey i Trimbur<sup>12</sup> oraz Azevedo i inni<sup>13</sup>.

<sup>11</sup>Harvey A.C., Jaeger A., 1993, *Detrending, Stylized Facts...*

<sup>12</sup>Harvey A.C., Trimbur T.M., 2003, *General Model-Based Filters...*

<sup>13</sup>Azevedo J.V.e, Koopman S.J., Rua A., 2003, *Tracking Growth and the Business Cycle: A Stochastic*

Po przedstawieniu strukturalnych modeli szeregów czasowych przedstawię parametry modelu tego typu, który wykorzystałem do estymacji równoczesnego wskaźnika koniunktury. Wybór  $m = 2$  i  $k = 6$  ten został uzasadniony przez Azevedo i inni<sup>14</sup> tym, że wówczas model ten ma właściwości zbliżone do filtru Baxtera-Kinga. Model ten równoległego wyestymowany był na  $N = 5$  zmiennych, a parametry były estymowane za pomocą metody największej wiarygodności. Częstotliwość  $\lambda_c$  wybrana została na poziomie  $\pi/48$ , który odpowiada wahaniom 96-miesięcznym i jest uzasadniony w literaturze przez Azevedo i innych<sup>15</sup> jako odpowiednia wartość do modelowania cyklu koniunkturalnego w Unii Europejskiej i Stanach Zjednoczonych. W modelu wskaźnika równoczesnego ograniczenie na parametr ładunku przy szeregu produkcji przemysłowej zostało narzucone tak, że  $\delta_1 = 1$ . W modelu indeksu wyprzedzającego podobne ograniczenie zostało nałożone na komponent cykliczny szeregu różnicy między rentownością 5-letnich obligacji skarbowych a 3-miesięcznym WIBORem. Model stochastycznego trendu (3.2)-(3.3) może być postrzegany jako filtr dolnoprzepustowy, czyli pozostawiający z całego zakresu wahań tylko te o najniższych częstotliwościach, czyli takie, których okres jest bardzo duży lub nieskończony. Model (3.4)-(3.5) jest filtrem średnioprzepustowym, gdyż z obserwowanego szeregu pozostawia tylko średnioczęstotliwościowe oscylacje. Harvey i Trimbur<sup>16</sup> określają model (3.1)-(3.4) jako uogólniony dolno- i średnioprzepustowy filtr Butterwortha rzędu  $(m, n)$ . Zgodnie z Harvey i Trimbur<sup>17</sup>, położenie maksimum funkcji wzmocnienia filtra średnioprzepustowego jest regulowane przez częstotliwość  $\lambda$ , natomiast kształt zależy od ustalenia parametru  $n$ . W oszacowanym modelu częstotliwość  $\lambda$  została ustalona na poziomie  $\pi/48$ , co odpowiada wahaniom 96-miesięcznym. Celem niniejszego podrozdziału nie jest jednak dokładne prezentowanie teorii filtracji, tylko zaprezentowanie modelu wynikającego z połączenia już istniejących.

Tablica 3.2: Oszacowania modelu wskaźnika równoczesnego

Parametry	Zmienne				
	Produkcja przemysłowa	Handel detaliczny	Import	Eksport	BWUK
$\delta_i$	1	0,434	1,742	1,728	0,733
$\sigma_i^2$	0,016	0,033	0,297	0,300	0,077
$\delta_i/\sigma_{\Delta y_i}$	43,434	14,597	42,092	46,637	13,496

*Źródło:* Opracowanie własne.

Parametr  $\rho$  został oszacowany na poziomie  $\rho = 0.487$ . Tablica 3.2 zawiera oszacowania modelu wskaźnika równoczesnego., natomiast komponent cykliczny przedstawiony jest na

*Common Cycle Model for the Euro Area*, Tinbergen Institute Discussion Paper, 2003-069/4.

<sup>14</sup>Azevedo J.V.e, Koopman S.J., Rua A., 2003, *Tracking Growth...*, str. 9.

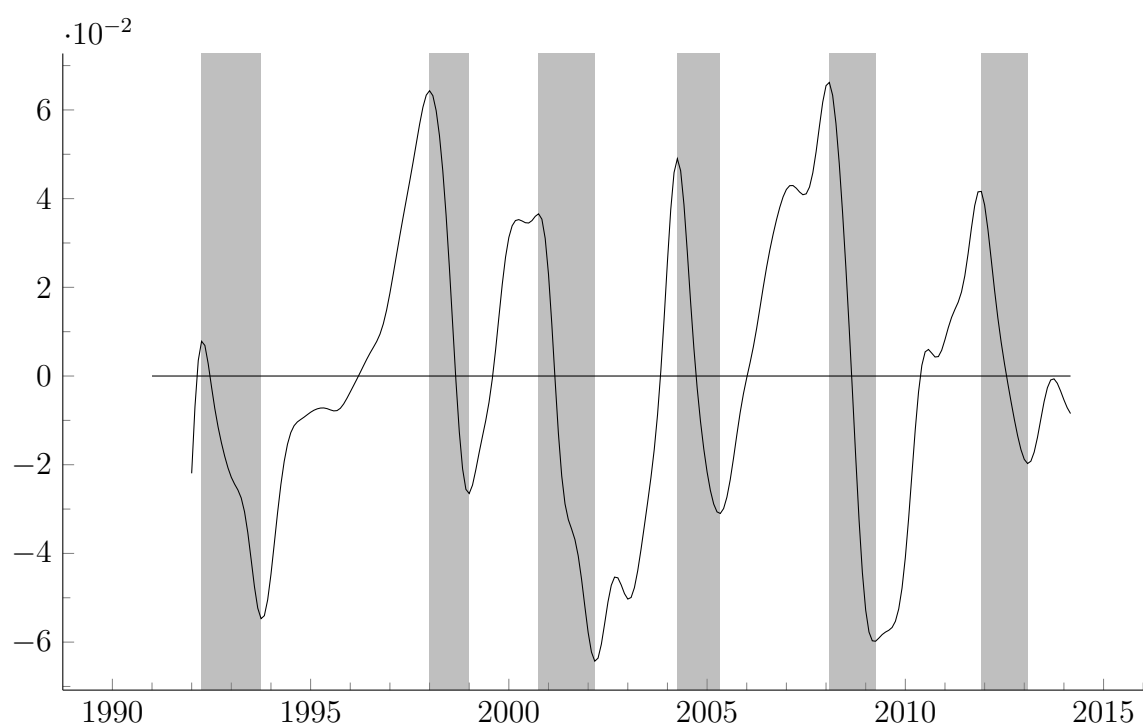
<sup>15</sup>Azevedo J.V.e, Koopman S.J., Rua A., 2003, *Tracking Growth...*, str. 9.

<sup>16</sup>Harvey A.C., Trimbur T.M., 2003, *General Model-Based Filters...*

<sup>17</sup>Harvey A.C., Trimbur T.M., 2003, *General Model-Based Filters...*, str. 246

rysunku 3.2. Zacienione obszary odpowiadają okresom od szczytu do dna koniunktury.

Rysunek 3.2: Komponent cykliczny równoczesnego wskaźnika koniunktury



Źródło: Opracowanie własne.

Graficzne porównania składnika cyklicznego z wynikami z filtru Hodricka-Prescotta, zmodyfikowanego filtru Hodricka-Prescotta i Christiano-Fitzgeralda, przedstawione w Skrzypczyńskim<sup>18</sup> i Skrzypczyńskim<sup>19</sup>, poświadczają duże podobieństwo przebiegu. W okresie, który może być porównany<sup>20</sup> stwierdzam podobieństwo za pomocą przypadających w tych samych okresach szczytów koniunktury w latach 1998, 2000 i 2004 między przedstawionymi w Skrzypczyńskim<sup>21</sup> a uzyskanym przeze mnie komponentem cyklicznym wskaźnika koniunktury. Wysoką zbieżność można zauważyć przy porównaniu przebiegu okresu ożywienia i ekspansji w latach 1999-2000, gdzie charakterystyczne dodatkowe lokalne maksimum jest obserwowane zarówno w przypadku wyników Skrzypczyńskiego<sup>22</sup> jak i moich. Adamowicz i inni<sup>23</sup> zaprezentowali szeroką analizę synchronizacji polskiego cyklu

<sup>18</sup>Skrzypczyński P., 2008, *Wahania aktywności gospodarczej w Polsce i strefie euro*, Materiały i Studia nr 227, Narodowy Bank Polski.

<sup>19</sup>Skrzypczyński P., 2010, *Metody spektralne w analizie cyklu koniunkturalnego gospodarki Polskiej*, Materiały i Studia nr 252, Narodowy Bank Polski.

<sup>20</sup>W pracach Skrzypczyńskiego przedmiotem analiz były dane za lata 1995-2007.

<sup>21</sup>Skrzypczyński P., 2010, *Metody spektralne w analizie cyklu...*, str. 142, 143, 146, 147 i 148.

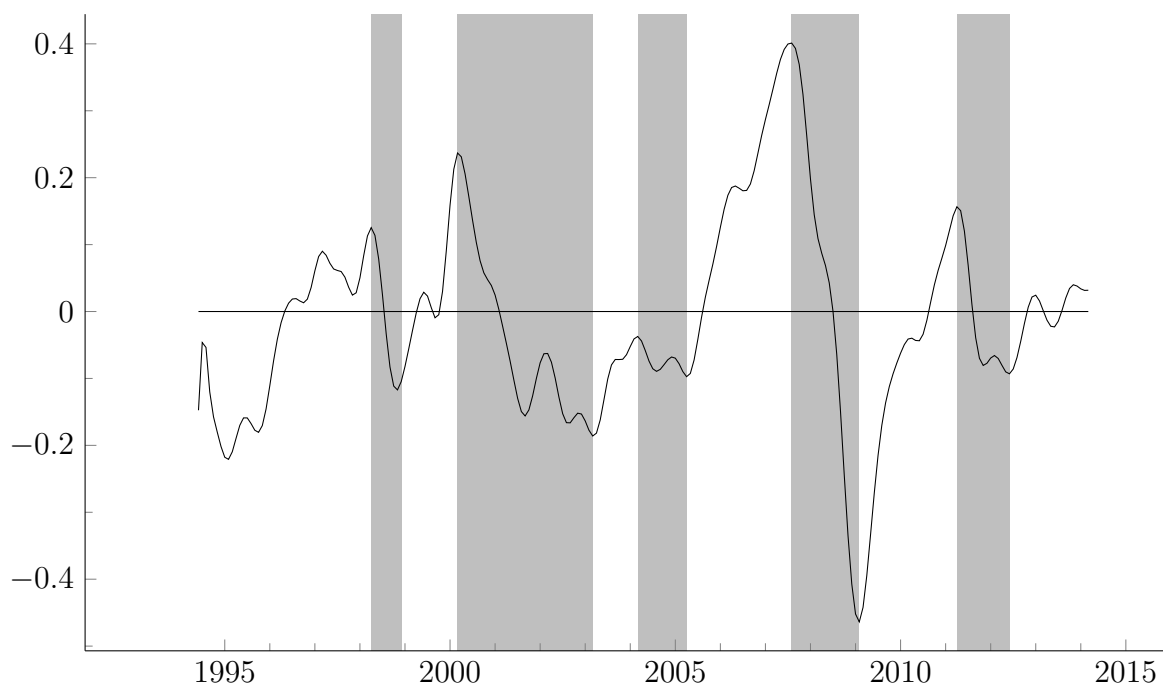
<sup>22</sup>Skrzypczyński P., 2010, *Metody spektralne w analizie cyklu...*, str. 148. Dodatkowe lokalne maksimum widoczne jest dla komponentów cyklicznych uzyskanych za pomocą filtrów Hodricka-Prescotta, modeli UCARIMA i SVAR.

<sup>23</sup>Adamowicz E., Dudek S., Pachnicki D., Walczyk K., 2009, *Synchronizacja cyklu koniunkturalnego polskiej gospodarki z krajami strefy euro w kontekście struktury tych gospodarek w Raport na temat pełnego uczestnictwa Rzeczypospolitej Polskiej w trzecim etapie Unii Gospodarczej i Walutowej*, Narodowy Bank Polski, Warszawa.

koniunkturalnego z krajami strefy euro. Wśród analizowanych przez nich szeregów znalazł się produkt krajowy brutto. Komponenty cykliczne wyznaczone za pomocą filtrów Christiano-Fitzgeralda i strukturalnego modelu czasowego z produktu krajowego brutto mają szczyty i dna w tych samych latach co wyliczone przeze mnie wskaźniki<sup>24</sup>. Podobny przebieg ma również eksperymentalny wskaźnik koniunktury GUS<sup>25</sup>, z tym, że zaprezentowany tam szereg charakteryzuje się większą gładkością.

Analogiczny model oszacowany został dla zmiennych wyprzedzających wymienionych w tablicy 3.1. Komponenty cykliczne oraz punkty zwrotne tych szeregów wykazują się wyprzedzającym charakterem względem wskaźnika równoczesnego. Jako punkt odniesienia w analizach przyjęty został szereg produkcji przemysłowej, który jest powszechnie wykorzystywany w tym celu w literaturze.

Rysunek 3.3: Komponent cykliczny wyprzedzającego wskaźnika koniunktury



Źródło: Opracowanie własne.

Przedstawiony powyżej model służyć może do uzyskania wskaźnika koniunktury z kilku szeregów o jednakowej, miesięcznej, częstotliwości. Zaletą tego podejścia jest brak potrzeby odfiltrowywania trendu z obserwowanych zmiennych, w przeciwieństwie np do metody stosowanej przez OECD. Niestety, wiele ważnych z punktu widzenia analizy koniunktury szeregów ekonomicznych, w tym najważniejszy szereg PKB, ma częstotliwość kwartalną, przez co nie mogą być dołączone do przedstawionego modelu stochastycznego cyklu i trendu. Literatura zawiera sposoby rozszacowania szeregów kwartalnych, które

<sup>24</sup>Szczyty obserwowane w latach 1998, 2000 i 2004 oraz dna z lat 1998, 2002, 2009 pokrywają się z punktami zmiany tendencji szeregów zaproponowanych przeze mnie.

<sup>25</sup>GUS, 2014, *Równoległy oraz wyprzedzający zagregowany wskaźnik koniunktury, zegar koniunktury (Opracowanie eksperymentalne)*, Główny Urząd Statystyczny, Warszawa.



wymienię w następnym podrozdziale i zaprezentuję moją propozycję wyznaczenia miesięcznych wartości produktu krajowego brutto.

Tablica 3.3: Oszacowania modelu wskaźnika wyprzedzającego

Parametry	Zmienne		
	Spread	WIG20	Oczekiwana produkcja sprzed. w najbl. 3 mies.
$\delta_i$	1	0,647	0,876
$\sigma_i^2$	0,332	0,026	0,122
$\delta_i/\sigma_{\Delta y_i}$	3,965	16,799	7,132

*Źródło:* Opracowanie własne.

## 3.2 Miesięczne PKB

Istnieje wiele metod dezagregacji zmiennych rocznych i kwartalnych. Artykuły Lismana i Sandee<sup>26</sup> oraz Boota i innych<sup>27</sup> prezentują metody służące do dezagregacji szeregów rocznych na kwortalne. Podobne metody zaproponowali również Zani<sup>28</sup>, Wei i Stram<sup>29</sup> czy Marcellino<sup>30</sup>. Wspólną cechą wspomnianych metod jest to, że nie wymagają dodatkowych informacji w postaci szeregów o wyższej częstotliwości i będących skorelowanych ze zmienną rozszacowywaną. Metody te określane są niekiedy mianem mechanicznej projekcji.

W artykule Chow i Lin<sup>31</sup> zostało przedstawione odmienne i bardzo popularne podejście do dezagregacji. W ramach tego podejścia, wykorzystującego pomocnicze zmienne o wyższej częstości, wymienić należy ponadto prace Rossi<sup>32</sup>, Littermana<sup>33</sup> i Di Fonzo<sup>34</sup>.

W klasie modeli przestrzeni stanów wyróżnić można dwa sposoby łączenia w modelu sze-

<sup>26</sup>Lisman J.H.C, Sandee J., 1964, *Derivation of Quarterly Figures from Annual Data*, Applied Statistics, 13, 87-90.

<sup>27</sup>Boot J.C.G., Feibes W., Lisman J.H.C., 1967, *Methods of Derivation of Quarterly Figures from Annual Data*, Journal of the Royal Statistical Society, Series C (Applied Statistics), 16(1), s. 65-75.

<sup>28</sup>Zani S., 1970, *Sui criteri di calcolo dei valori trimestrali di tendenza degli aggregate della contabilità nazionale*, Studi di Ricerche, Facolta di Economia e Commercio Universta degli Studi di Parma VII, s. 287-349.

<sup>29</sup>Wei W.W.S., Stram D.O., 1990, *Disaggregation of Time Series Models*, Journal of the Royal Statistical Society. Series B (Methodological), 52(3), s. 453-467.

<sup>30</sup>Marcellino M., 1998, *Temporal Disaggregation, Missing Observations, Outliers, and Forecasting: A Unifying Non-Model-Based Procedure*, Advances in Econometrics, 13, s. 181-202.

<sup>31</sup>Chow G.C., Lin A., 1971, *Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series*, The Review of Economics and Statistics, 53, s. 372-375.

<sup>32</sup>Rossi N., 1982, *A Note on the Estimation of Disaggregate Time Series When The Aggregate is Known*, The Review of Economics and Statistics, 64(4), s. 695-696.

<sup>33</sup>Litterman R.B., 1983, *A Random Walk, Markov Model for Distribution of Time Series*, Journal of Business and Economics Statistics, 1, s. 169-173.

<sup>34</sup>Di Fonzo T., 1990, *The Estimation of M Disaggregate Time Series when Contemporaneous and Temporal Aggregates Are Known*, Review of Economics and Statistics, s. 178-182.

regów o różnych częstotliwościach. Pierwszy jest opisany w Harvey<sup>35</sup> i można go rozumieć jako interpolację zmiennej kwartalnej przyz uciu ymiennej kumulującej. Drugi sposób, opisany w Mariano i Murasawa<sup>36</sup> jest rozszerzeniem modelu Stocka i Watsona<sup>37</sup> o zmienną kwartalną. Dodatkowe rozszerzenie modelu Mariano i Murasawy o zmienną kumulującą zaproponował Nunes<sup>38</sup>.

Moja propozycja jest rozszerzeniem istniejącego modelu wspólnego stochastycznego cyklu, wykorzystanego do wyliczenia wskaźnika koniunktury w poprzednim podrozdziale, o zmienną kumulującą. Rozważmy przedstawiony w poprzednim podrozdziale model wspólnego stochastycznego cyklu z pewnymi rozszerzeniami. Niech, podobnie jak poprzednio,  $y_{it}$  będzie  $i$ -tą obserwowaną zmienną, która modelowana jest następująco:

$$y_{it} = \mu_{it} + \delta_i \{ \cos(\xi_i \lambda) \psi_{n,t} + \sin(\xi_i \lambda) \psi_{n,t}^* \} + \varepsilon_{it}, \quad (3.6)$$

$$\mu_{it} = \mu_{it-1} + \nu_{it-1}, \quad (3.7)$$

$$\nu_{it} = \nu_{it-1} + \zeta_{it}, \quad (3.8)$$

$$\begin{bmatrix} \psi_{j,t} \\ \psi_{j,t}^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{j,t-1} \\ \psi_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \psi_{j-1,t-1} \\ \psi_{j-1,t-1}^* \end{bmatrix}, \quad (3.9)$$

$$\begin{bmatrix} \psi_{1,t} \\ \psi_{1,t}^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{1,t-1} \\ \psi_{1,t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad (3.10)$$

dla  $j = 2, \dots, n$ , gdzie  $t$  jest indeksem czasu  $t = 1, 2, \dots, T$ , natomiast  $i$  indeksem dla zmiennych  $i = 1, 2, \dots, N$ . Zgodnie z Azevedo i innymi<sup>39</sup>, parametr  $\xi_i$  interpretuje się jako przesunięcie fazowe  $i$ -tej zmiennej względem cyklu referencyjnego, przy czym  $\xi_i > 0$  ( $\xi_i < 0$ ) oznacza wyprzedzanie (opóźnianie) względem cyklu szeregu referencyjnego. Zagadnienie dodania do modelu zmiennej o niższej częstotliwości wiąże się z dodaniem kolejnej zmiennej stanu, która jest nieobserwowalnym odpowiednikiem tej zmiennej. Zmienna kwartalna jest obserwowana tylko w co trzecim miesiącu, więc niezbędne jest wprowadzenie zmian do modelu. W efekcie braku informacji o wartościach zmiennej kwartalnej w niektórych okresach czasu niezbędne jest wprowadzenie zmian do modelu (3.6)-(3.10).

Niech  $y_q$  będzie zmienną kwartalną, która ma zostać włączona do miesięcznego modelu określonego równaniami (3.6)-(3.10). Zmienna kwartalna określona jest jedynie w punktach czasu  $t = 3s$ , dla  $s = 1, 2, \dots, \lfloor \frac{T}{3} \rfloor$ . Niech  $y_q^\dagger$  będzie zmienną zdefiniowaną dla  $t = 1, 2, \dots, T$  tak, że  $y_{q,3s}^\dagger = y_{q,3s}$  i nieobserwowaną w punktach czasu  $3s - 2$  i  $3s - 1$ . Zmienna  $y_{q,t}^\dagger$  nie jest sprzeczna z modelem (3.6)-(3.10), gdyż jest określona dla tych samych

<sup>35</sup>Harvey A. C., 1989, *Forecasting, Structural Time Series...*

<sup>36</sup>Mariano R.S., Murasawa Y., *A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series*, PIER Working Paper 02-014, Philadelphia 2002.

<sup>37</sup>Stock J.H., Watson M.W., *A Probability Model of The Coincident Economic Indicators*, Opublikowane w: G. Moore and K. Lahiri, editors. *The Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge 1990.

<sup>38</sup>Nunes L.C., 2005, *Nowcasting Quarterly GDP Growth in a Monthly Coincident Indicator Model*, Journal of Forecasting, 24, s. 575-592.

<sup>39</sup>Azevedo J.V.e, Koopman S.J., Rua A., 2003, *Tracking Growth...*, str. 5.

punktów w czasie,  $t = 1, 2, \dots, T$ . Równanie pomiarowe dla zmiennej  $y_q^\dagger$  ma postać:

$$y_{q,\delta t} = \mu_{q,t} + \delta_q \psi_{n,t} + \varepsilon_{q,t}. \quad (3.11)$$

Równanie (3.11) prawdziwe jest tylko dla punktów w czasie, w których dostępne są obserwacje zmiennej kwartalnej, to jest w ostatnim miesiącu każdego kwartału. W przypadku nałożenia ograniczenia  $\delta_q = 1$ , wspólny komponent cykliczny będzie miał amplitudę równą amplitudzie składnika cyklicznego zmiennej kwartalnej. Konsekwentnie, parametry  $\delta_i$  przy zmiennych miesięcznych będą estymowane. Nieobserwowany odpowiednik zmiennej kwartalnej wyliczany jest ze wzoru

$$y_t^\dagger = \mu_{m,t}^q + \psi_{n,t} + \varepsilon_t, \quad (3.12)$$

czyli jako suma komponentu trendu, składnika cyklicznego i składnika nieregularnego.

Metoda pozwala na interpolację brakujących wartości składnika cyklicznego zmiennej kwartalnej na podstawie zmiennych miesięcznych. Składnik cykliczny jest szeregiem odchyleń od trendu w określonym zakresie częstotliwości. W analizach koniunktury na podstawie takiego wskaźnika możliwe jest wyznaczenie punktów zwrotnych cyklu, w których następuje zmiana średniookresowych tendencji. Punkty zwrotne, nazywane też punktami zmiany tendencji, są lokalnymi ekstremami szeregu reprezentującego koniunkturę. Na wyznaczenie punktów zwrotnych nie ma wpływu monotoniczna transformacja szeregu. Fakt ten jest wykorzystany w Azevedo i inni<sup>40</sup>, gdzie składnik cykliczny został potraktowany jako zmienna stanu, przez co analiza koniunktury w strefie euro została dokonana na interpolacji składnika cyklicznego PKB. Wyliczenie miesięcznego PKB wymaga dodania do modelu ograniczenia sumowalności wartości miesięcznych do sum kwartalnych za pomocą zmiennej kumulującej, którą prezentuję poniżej.

Niech  $y_{pkb,3s}$  będzie kwartalnym szeregiem produktu krajowego brutto obserwowanym tylko dla miesięcy kończących kwartał. Zgodnie z powyższym zdefiniowana musi zostać zmienna  $y_{pkb,t}^\dagger$ , która jest nieobserwowanym odpowiednikiem  $y_{pkb,3s}$  dla wszystkich okresów czasu  $t = 1, 2, \dots, T$ . Zagadnienie miesięcznego PKB polega natomiast na wyliczeniu  $y_{pkb,t}^\dagger$ , takiego, że

$$y_{pkb,3s} = y_{pkb,3s-2}^\dagger + y_{pkb,3s-1}^\dagger + y_{pkb,3s}^\dagger. \quad (3.13)$$

Rozwiązaniem problemu sumowalności miesięcznego PKB do wartości kwartalnych jest użycie zmiennej kumulującej, tak jak przedstawiono to w Harvey'u<sup>41</sup>. Zmienna kumulująca,  $y_t^f$ , dla zmiennej przepływu określona jest w sposób, że dla pierwszego miesiąca kwartału równa jest nieobserwowanemu miesięcznemu PKB,  $y_{pkb,t}^\dagger$ , natomiast dla kolejnych miesięcy równa jest sumie  $y_{t-1}^f$ , swojej przeszłej wartości, i  $y_{pkb,t}^\dagger$ , czyli miesięcznego PKB w tym okresie. Rekurencyjna zależność dla zmiennej kumulującej może być zapisana

<sup>40</sup>Azevedo J.V.e, Koopman S.J., Rua A., 2003, *Tracking Growth...*

<sup>41</sup>Harvey A. C., 1989, *Forecasting, Structural Time Series...*, str. 313.

w postaci

$$y_{3(s-1)+r}^f = \varphi_{3(s-1)+r} y_{3(s-1)+r-1}^f + y_{3(s-1)+r}^\dagger, \quad (3.14)$$

gdzie  $r = 1, 2, 3$ , natomiast  $\varphi_{3(s-1)+r}$  zdefiniowane jest jako

$$\varphi_{3(s-1)+r} = \begin{cases} 0 & \text{dla } r = 1, \\ 1 & \text{dla } r = 2, 3. \end{cases} \quad (3.15)$$

W przypadku danych kwartalnych w miesięcznym modelu  $\varphi_{\delta(t-1)+s}$  jest indykátorem, który przyjmuje wartość zero dla pierwszego miesiąca w kwartale.

Zastosowany model miesięcznego PKB jest oparty na równaniach (3.6)-(3.10), (3.13)-(3.14). Wektor zmiennych obserwowanych ma postać

$$y = \begin{bmatrix} \text{Produkt krajowy brutto} \\ \text{Produkcja przemysłowa} \\ \text{Import} \\ \text{Eksport} \\ \text{Handel detaliczny} \\ \text{WIG20} \\ \text{Rentowność 5-letnich obligacji skarbu. – 3-mies. WIBOR} \\ \text{Bieżący wskaźnik ufności konsumenckiej} \end{bmatrix}. \quad (3.16)$$

Ze względu na identyfikację parametrów musi zostać nałożone ograniczenie  $\delta_{pkb} = 1$ , przez co wspólny komponent cykliczny będzie miał amplitudę równą amplitudzie składnika cyklicznego zmiennej kwartalnej. Zastosowany został model stochastycznego cyklu rzędu 6 oraz model lokalnego liniowego trendu. Wybór  $\lambda = \frac{\pi}{48}$ ,  $m = 2$  oraz  $n = 6$  uzasadniony jest analogicznie jak w podrozdziale 3.1. W modelu rozszacowywany był odsezonowany szereg produktu krajowego brutto w cenach stałych przedstawiony jako indeks o bazie w roku 1995 publikowany zgodnie z metodologią ESA95. Oszacowania parametrów modelu znajdują się w tablicy 3.4.

Komponent cykliczny miesięcznego PKB ma bardzo zbliżony przebieg do komponentu cyklicznego z modelu wskaźnika równoczesnego. Wnioski wyciągnięte z porównania wahań średniookresowych wskaźnika koniunktury pozostają niezmiennione dla komponentu średniookresowego miesięcznego PKB.

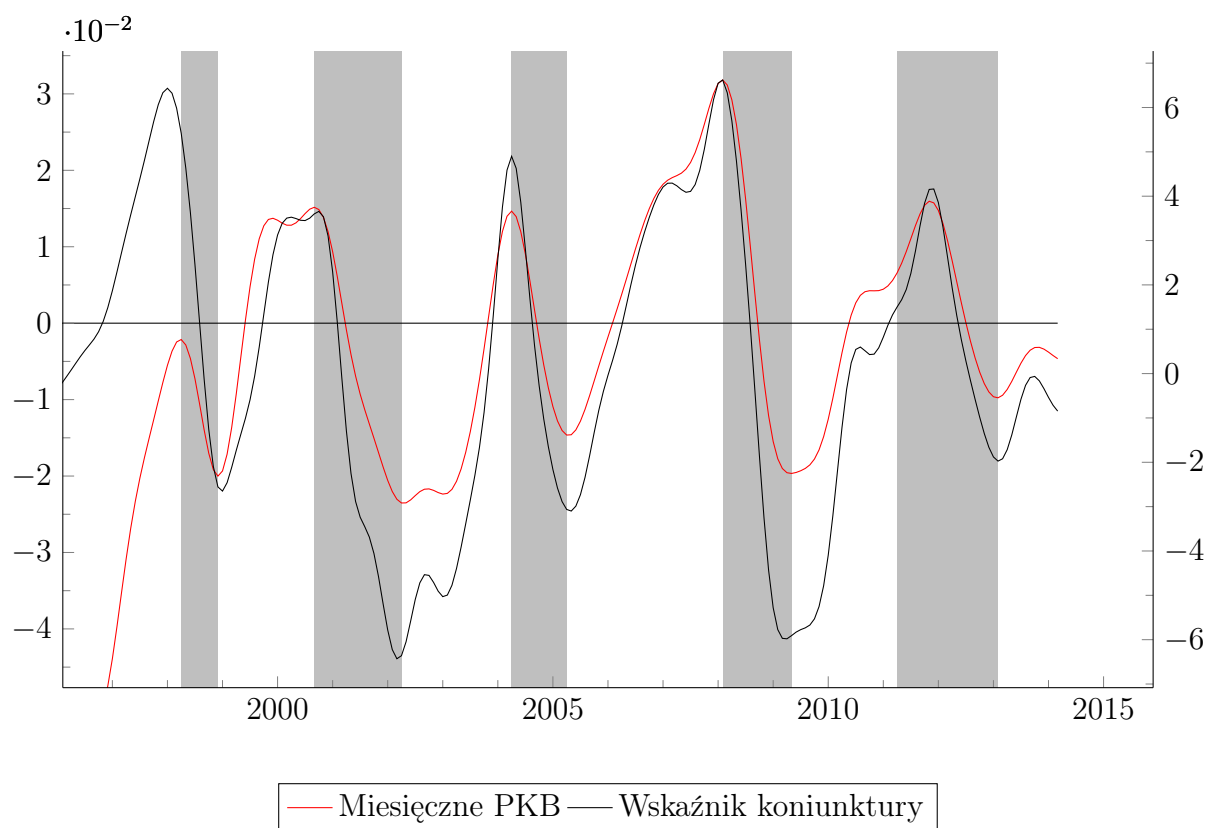
Zgodnie z terminologią NBER punktem zwrotnym lub punktem zmiany tendencji (*turning point*) jest pojedynczy okres czasu, w którym następuje zmiana pomiędzy okresami wzrostu i kurczenia się gospodarki. Pikiem lub szczytem koniunktury (*peak*) jest punkt czasu, w którym gospodarka przestaje się rozwijać i zaczyna się kurczyć. Odwrotnym przypadkiem jest dno (*trough*), które jest okresem granicznym między fazą kurczenia się a fazą wzrostu. Zgodnie z definicją cyklu koniunkturalnego okres ekspansji kończy się szczytem i po nim zaczyna się okres dekonunktury, czyli kurczenia się zagregowanej

Tablica 3.4: Oszacowania modelu miesięcznego PKB

Zmienna $y_i$	Ładunek $\delta_i$	Przesunięcie $\xi_i$	Wariancja $\varepsilon_i$ $\sigma_{\varepsilon_i}^2$
PKB, ceny stałe (1995=100)	1	–	0,007
Produkcja przemysłowa	2,502	0,129	0,015
Import	2,601	-0,130	0,033
Eksport	2,510	0,396	0,028
Handel detaliczny	1,253	0,294	0,025
WIG20	1,970	-0,080	0,110
Spread	0,355	-0,352	0,090
BWUK	1,469	0,197	0,080

Źródło: Opracowanie własne.

Rysunek 3.4: Komponenty cykliczne miesięcznego PKB i równoczesnego wskaźnika koniunktury



Źródło: Opracowanie własne.

produkcji. W momencie, gdy ta skurczy się poniżej poziomu wartości z długookresowej ścieżki wzrostu, wówczas mamy do czynienia z okresem recesji. Faza recesji kończy się dnem, czyli lokalnym średniookresowym minimum wielkości gospodarki. Okres ożywienia po nim następujący, trwa do momentu znalezienia się ponad długookresowym trendem, od którego rozpoczyna się stan ekspansji trwający aż do szczytu.

Do wyznaczenia punktów zwrotnych użyte zostały procedury Bry-Boschan<sup>42</sup> dla danych miesięcznych oraz Hardinga-Pagana<sup>43</sup> dla kwartalnego PKB. Wykorzystane zostały implementacje tych procedur w pakiecie Grocer 1.6, autorstwa Dubois i Michaux<sup>44</sup>, środowiska SciLab. Tablica 3.5 prezentuje ustalone punkty zmiany tendencji dla oszacowanych wskaźników, kwartalnego PKB oraz punktami zwrotnymi polskiej gospodarki opublikowanymi przez OECD<sup>45</sup>. Procedury Bry-Boschan i Hardinga-Pagana są formalnymi procedurami wyznaczania punktów zwrotnych, co uzasadnia ich użycie. Chauvet i Piger<sup>46</sup> przytoczyli zalety formalnych procedur określania punktów zwrotnych, w tym metody Bry-Boschan. Analiza otrzymanych wyników pozwala stwierdzić wysoką zbieżność punktów zwrotnych zaproponowanych miesięcznych wskaźników równoczesnych z punktami zmiany tendencji kwartalnego szeregu PKB. Uzasadnia to użycie indeksu miesięcznego PKB oraz jego składnika cyklicznego do analiz korelacji i przyczynowości między szeregami popytu na energię elektryczną a stanem gospodarki.

Recesja związana z amerykańskim rynkiem kredytów hipotecznych znajduje odzwierciedlenie w polskim cyklu koniunkturalnym. Pik krajowego produktu brutto przypada na drugi kwartał 2008, podczas gdy wskaźniki miesięczne wskazują luty 2008 roku jako punkt początkowy dekonunktury. Największa niezgodność zaproponowanych indeksów koniunktury z kwartalnym PKB dotyczy kolejnego punktu zwrotnego, który miał miejsce w kwietniu lub maju 2009, zgodnie z odpowiednio miesięcznym lub wskaźnikiem koniunktury. Na podstawie analizy kwartalnego PKB polska gospodarka osiągnęła ten punkt zmiany tendencji w cyklu dopiero w pierwszym kwartale 2010 roku. W 2011 roku miał miejsce kolejny szczyt koniunktury kolejnego cyklu, choć medialnie nośny „kryzys” utrzymywał się jeszcze przez jakiś czas. Okres słabszej koniunktury po 2011 roku zakończył się w lutym 2013, zgodnie z datowaniem cykli za pomocą miesięcznego PKB i wskaźnika równoczesnego.

Otrzymane wyniki dla okresu 1996-2009 są zgodne z przedstawionymi w Gradzewicz i inni (2010). Cztery z siedmiu punktów zwrotnych są zgodne co do kwartału, natomiast pozostałe przypadają na kwartały sąsiednie, wobec czego stwierdzić należy dość dużą

<sup>42</sup>Bry G., Boschan Ch., 1971, *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, Columbia University Press.

<sup>43</sup>Harding D., Pagan A., 2002, *Dissecting the Cycle: a Methodological Investigation*, Journal of Monetary Economics, 49, s. 365-381.

<sup>44</sup><http://dubois.ensae.net/grocer.html> Dostęp 21 maja 2015.

<sup>45</sup><http://www.oecd.org/std/leading-indicators/CLI-components-and-turning-points.pdf>. Dostęp 6 lipca 2014.

<sup>46</sup>Chauvet M., Piger J.M., 2005, *A Comparison of the Real-Time Performance of Business Cycle Dating Methods*, Working Papers 2005-021, Federal Reserve Bank of St. Louis.

Tablica 3.5: Porównanie punktów zwrotnych wskaźników i PKB

MPKB		Wskaźniki		OECD		Kwartalne
		COIN	LEI	Opublikowane	LEI	PKB
1998.04	S	-3	0	-1	-9	1998q2
1998.12	D	+1	-1	-	-8	1999q1
2000.09	S	+1	-6	-	-18	2000q1
2002.04	D	-1	+11	+8	-10	2002q4
2004.04	S	0	-1	0	-1	2004q2
2005.04	D	0	0	+2	0	2005q2
2008.02	S	0	-6	0	-12	2008q2
2009.05	D	-1	-3	+3	-7	2010q1
2011.04	S	+8	0	+7	-8	2011q3
2013.02	D	0	-8	-	-3	-

W tabeli przyjęto oznaczenia: S - szczyt, D - dno koniunktury. MPKB - oznacza miesięczne PKB, COIN - wskaźnik koniunktury, LEI - indeks wyprzedzający. *Źródło:* Opracowanie własne.

zgodność oszacowań. Zauważalne na rysunkach 3.3 i 3.4 krótsze fazy spowolnienia i recesji niż ożywienia i rozkwitu są potwierdzone przez analizę punktów zwrotnych, co zostało dostrzeżone także w Gradzewiczu i innych<sup>47</sup>. Do formalnego porównania punktów zwrotnych wykorzystałem statystykę  $CC_{corr}$  przedstawioną przez Artisa i innych<sup>48</sup>. Wartość uzyskanej statystyki równa 80,13% może być zinterpretowana jako istnienie silnej korelacji między komponentem cyklicznym równoczesnego wskaźnika koniunktury a komponentem cyklicznym kwartalnego PKB<sup>49</sup>.

### 3.3 Dekompozycja szeregów energii elektrycznej

Jednym ze sposobów weryfikacji hipotezy o wpływie cyklu koniunkturalnego na popyt na energię elektryczną może być wyodrębnienie komponentów cyklicznych z danych go opisujących i porównanie ich z cyklem koniunkturalnym. Składowa średniookresowa wyznaczyłem za pomocą filtru średnio-przepustowego Christiano-Fitzgeralda, który był stosowany w analizach polskiego cyklu koniunkturalnego, na przykład przez Skrzypczyńskiego<sup>50</sup> czy Gradzewicza i innych<sup>51</sup>. Drugim narzędziem wykorzystanym do sprawdzenia czy szeregi energii elektrycznej zawierają wahania w paśmie przypisywanemu wahaniom koniunkturalnym

<sup>47</sup>Gradzewicz M., Growiec J., Hagemeyer J., Popowski P., 2010, *Cykl koniunkturalny w Polsce – wnioski z analizy spektralnej*, Bank i Kredyt, 41(5), s. 41-76.

<sup>48</sup>Artis M.J., Kontolenis Z.G., Osborn D.R., 1997, *Business Cycles for G7 and European Countries*, Journal of Business, 70(2), s. 249-279.

<sup>49</sup>Dla porównania komponentu cyklicznego miesięcznego PKB z komponentem cyklicznym kwartalnego PKB uzyskałem tą samą wartość statystyki  $CC_{corr}$ .

<sup>50</sup>Skrzypczyński P., 2010, *Metody spektralne...*

<sup>51</sup>Gradzewicz M., Growiec J., Hagemeyer J., Popowski P., 2010, *Cykl koniunkturalny...*

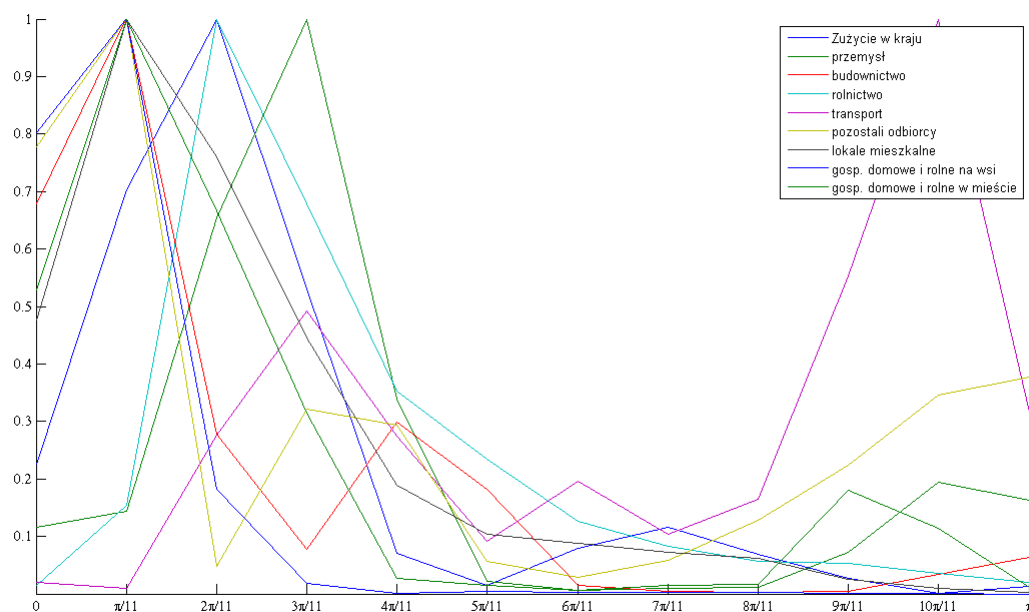
ralnym była analiza spektralna.

Szeregi czasowe o popycie na energię elektryczną dla Polski zawarte są w seriach publikacji Agencji Rynku Energii<sup>52</sup>. Tabele z danymi w nich zawartymi, nie są dostępne w wersji elektronicznej, więc niezbędne było ich własnoręczne przepisanie. Powstała w ten sposób baza danych zawierała 2 szeregi miesięczne, 9 rocznych i 4 z łamaną częstotliwością. Do analizy tych ostatnich niezbędne było niestandardowe podejście.

### 3.3.1 Dane roczne

Częstotliwość roczną mają szeregi całkowitego zużycia energii elektrycznej w kraju oraz w podziale na gałęzie gospodarki: przemysł, budownictwo, transport, pozostali odbiorcy. Dodatkowo analizowane były szeregi zużycia energii w lokalach mieszkalnych oraz w gospodarstwach domowych i rolnych w podziale na miasto i wieś. Dane pokrywają okres od 1990 do 2012 roku, czyli dostępne były 23 obserwacje.

Rysunek 3.5: Znormalizowane funkcje gęstości widmowej szeregów rocznych



Źródło: Opracowanie własne.

Rysunek 3.5 przedstawia znormalizowane funkcje gęstości widmowej dla szeregów rocznych. Zmienne zostały pozbawione liniowego trendu oraz przemnożone przez okno Blackmana-

<sup>52</sup>ARE, *Informacja statystyczna o energii elektrycznej*, Agencja Rynku Energii, Warszawa. ARE, *Statystyka Elektroenergetyki Polskiej*, Agencja Rynku Energii, Warszawa. ARE, *Sytuacja energetyczna w Polsce - krajowy bilans energii*, Agencja Rynku Energii, Warszawa. ARE, *Sytuacja w elektroenergetyce*, Agencja Rynku Energii, Warszawa.



Harrisa, co miało na celu rozwiązanie problemu wycieku mocy<sup>53</sup>. Maksima funkcji gęstości dla zużycia ogółem i w przemyśle przypadają na wahania powtarzające się odpowiednio co około 8 i 6 lat. Okresy wpadają w definicyjny zakres wahań koniunkturalnych, co może być argumentem potwierdzającym hipotezę o obecności wahań koniunkturalnych w szeregach popytu na energię elektryczną. W zużyciu energii elektrycznej przez podmioty, których działalność gospodarcza jest klasyfikowana jako transport wykazują silne wahania powtarzające się co 2,2 roku. Dla pozostałych szeregów dominujące fluktuacje mają okres zbliżony do górnej granicy definicji cyklu koniunkturalnego.

Tablica 3.6: Wyniki analizy korelacji krzyżowych komponentów cyklicznych

Zmienna	Współczynnik korelacji	Opóźnienie
Zużycie w kraju	0,798	0
Przemysł	0,308	0
Budownictwo	0,212	0
Rolnictwo	0,302	-1
Transport	0,443	0
Pozostali odbiorcy	0,355	0
Lokale mieszkalne	-0,355	0
Gosp. domowe i rolne na wsi	-0,270	0
Gosp. domowe i rolne w mieście	-0,202	1

Dodatnia wartość opóźnienia oznacza, że komponent cykliczny PKB jest opóźniony o daną liczbę okresów w stosunku do analizowanego szeregu. *Źródło: Opracowanie własne.*

Przy użyciu filtra Christiano-Fitzgeralda wyznaczone zostały składowe średniookresowe rocznych szeregów energii elektrycznej i PKB. Tablica 3.6 przedstawia wyniki badania korelacji krzyżowej<sup>54</sup> komponentów cyklicznych szeregów z cyklem produktu krajowego brutto. Odnotować należy silną korelację zużycia krajowego z PKB oraz umiarkowaną dodatnią korelację przemysłu a PKB. Komponenty cykliczne wszystkich zmiennych poza rolnictwem i miejskich gospodarstw domowych są równoczesne ze składową cykliczną PKB. Interesującym wnioskiem jest antycykliczność zapotrzebowania na energię elektryczną w lokalach mieszkalnych. Dane roczne w ogólności wykazują wysoką korelację z szeregiem produktu krajowego brutto, dlatego w następnym kroku sprawdzę istnienie korelacji dla danych miesięcznych.

### 3.3.2 Dane miesięczne

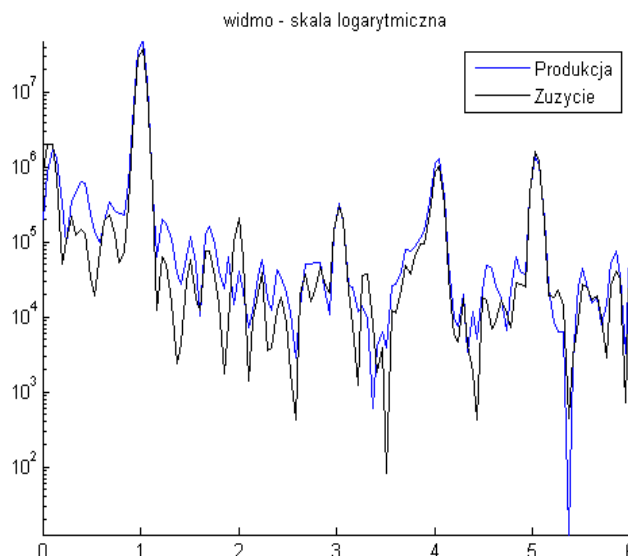
Analogiczne postępowanie zostało przeprowadzone dla danych miesięcznych. W tym podrozdziale wykorzystałem szeregi produkcji energii elektrycznej w kraju oraz zużycia kra-

<sup>53</sup>Problem wycieku mocy został omówiony w podrozdziale 1.2.

<sup>54</sup>Korelacja krzyżowa jest narzędziem statystycznym służącym do analizowania korelacji między dwoma szeregami czasowymi z uwzględnieniem wzajemnych przesunięć w czasie.

jowego. Dane pokrywały okres od stycznia 1993 r. do sierpnia 2013 roku. Gęstość spektralna miesięcznych szeregów produkcji energii elektrycznej ogółem oraz zużycia krajowego ujawniają mocną składową sezonową. Dla obydwu zmiennych występują ponadto składowe o częstotliwościach z zakresu cyklu koniunkturalnego. W tym zakresie produkcja energii ma maksimum widma dla częstości odpowiadającej wahaniom powtarzającym się co około 30 miesięcy, natomiast w szeregu zużycia energii elektrycznej występują istotne wahania o okresie 41 miesięcy.

Rysunek 3.6: Funkcje gęstości spektralnych zmiennych miesięcznych



Źródło: Opracowanie własne.

Porównałem także składowe średniookresowe z komponentem cyklicznym koniunktury za pomocą korelacji krzyżowej. Do wyznaczenia składowych cyklicznych wykorzystałem, tak jak w poprzednim podrozdziale, filtr Christiano-Fitzgeralda. Ze względu na miesięczną częstotliwość danych przyjąłem za referencyjne trzy szeregi: wyliczone w podrozdziałach 3.1 i 3.2 wskaźniki koniunktury i miesięczne PKB oraz miesięczny indeks produkcji przemysłowej. Wyniki są zaprezentowane w tabelicy 3.7 i na ich podstawie stwierdzić należy wyprzedzający charakter szeregów energii elektrycznej.

### 3.3.3 Dane z łamaną częstotliwością

Część danych publikowanych przez Agencję Rynku Energii ma zmienną w czasie częstotliwość. Dla szeregów przedstawiających wielkości energii elektrycznej przekazanej odbiorcom finalnym, ogółem oraz w podziale według napięcia<sup>55</sup>, od stycznia 1993 do grudnia 2002 dostępne są dane miesięczne, w okresie 2003-2007 - dane kwartalne, natomiast od

<sup>55</sup>W publikowanych statystykach Agencji Rynku Energii istnieje podział na sieci wysokiego napięcia (WN), średniego napięcia (SN) i niskiego napięcia (nN)

Tablica 3.7: Wyniki analizy korelacji krzyżowej między miesięcznymi szeregami energii elektrycznej a stanem gospodarki.

Zmienna	Wskaźnik COIN		Miesięczne PKB		Prod. przemysłowa	
	Korelacja	Opóźnienie	Korelacja	Opóźnienie	Korelacja	Opóźnienie
Produkcja energii	0,609	4	0,183	6	0,489	2
Zużycie energii	0,700	2	0,129	3	0,452	0

Dodatknie opóźnienie oznacza, że komponent cykliczny zmiennej wyprzedza koniunkturę.  
*Źródło:* Opracowanie własne.

2008 dane publikowane są z częstotliwością roczną. Do wyznaczenia gęstości widmowej szeregów z łamaną częstotliwością zastosowałem transformację Fouriera dla danych nierówno rozłożonych w czasie. Zmienne, które analizuję w tym podrozdziale mają charakter przepływu, więc wartość za dany rok jest sumą wartości zmiennej z miesięcy tego roku<sup>56</sup>.

Do wyznaczenia transformaty Fouriera niezbędne było posiadanie wartości dla każdego miesiąca kończącego kwartał w latach 2003-2007 oraz dla każdego grudnia od 2008 do 2012 roku. Ich przybliżone wartości oszacowałem na podstawie próby danych miesięcznych za lata 1993-2002. Dla każdego ostatniego miesiąca kwartału wyznaczyłem jego średni udział w wartości kwartalnej oraz krańce 90% przedziału ufności dla tej wartości. Dla grudnia obliczyłem przeciętny udział wartości miesięcznej w danych rocznych. W związku z tym, dla każdego ostatniego miesiąca kwartału od 2003 do 2007 oraz dla każdego grudnia w latach 2008-2012 dysponowałem 3 wartościami. Przy przyjętych przybliżeniach wartość szeregu w każdym wspomnianym punkcie mogła być równa jednej z trzech alternatyw, co przekładało się na  $3^{25}$  możliwych realizacji zmiennej.

Ze względu na czas estymacji analizę ograniczono do 30000 możliwych realizacji, które były wybierane w sposób losowy<sup>57</sup>. Dzięki temu oszacowane zostało pasmo, w którym powinno znajdować się teoretyczne widmo dla szeregu bez braków obserwacji. Przedstawione jest ono na rysunku 3.7. Gęstości spektralne obliczone zostały przy użyciu szybkiej transformacji Fouriera dla nierównomiernie rozłożonych w czasie danych. Wykorzystałem pakiet dla programu MatLab umożliwiający wyliczenie transformacji Fouriera dla takich danych zaprogramowany przez Greengarda i Lee i opisany w Greengard i Lee<sup>58</sup> oraz Lee i Greengard<sup>59</sup>.

Przebieg funkcji gęstości widmowej jest stabilny przy różnych realizacjach wartości mie-

<sup>56</sup>Inaczej niż w zmiennych stanu, których wartość jest średnią z okresu lub wartością za ostatni dzień okresu jak na przykład dla kursów walutowych.

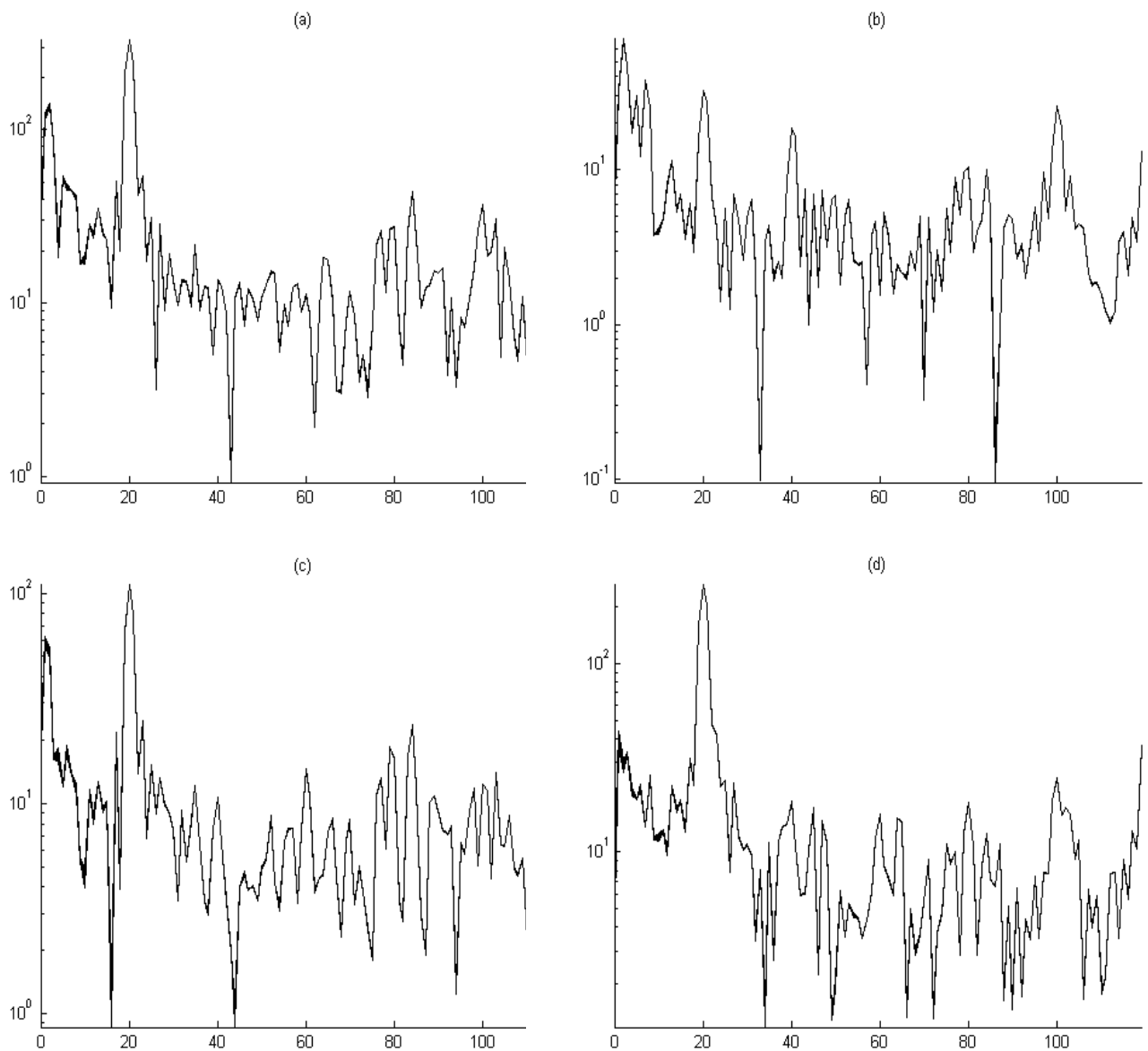
<sup>57</sup>W każdej iteracji do każdej z wartości alternatywnych dla każdego miesiąca kończącego kwartał w latach 2003-2007 i każdego grudnia z lat 2008-2012 przypisałem równe prawdopodobieństwa wyboru.

<sup>58</sup>Greengard L.F., Lee J.-Y., 2004, *Accelerating the Nonuniform Fast Fourier Transform*, SIAM Review 46(3), 443-454.

<sup>59</sup>Lee J.-Y., Greengard L.F., 2005, *The Type 3 Nonuniform FFT and its Applications*, Journal of Computational Physics, 206(1), 1-5.

sięczych w okresie 2003-2012, wobec czego wnioski powinny być prawidłowe. Lokalne maksima funkcji mocy przypadają dla częstości 2, 5, 13, 17, które odpowiadają wahaniom o okresach około 10, 4 lat, 18- oraz 14-miesięcznym. Pik dla wartości 20 odpowiada miesięcznej sezonowości. Wahania o częstotliwości 10-letniej są pozostałością po usuniętych trendzie. Obecność wahań powtarzających się co 4 lata i co 1,5 roku świadczy o występowaniu wahań zgodnych z cyklem koniunkturalnym.

Rysunek 3.7: Funkcje gęstości widmowej szeregów energii elektrycznej dostarczonej odbiorcom (a) ogółem, (b) na WN, (c) na SN i (d) na nN



Źródło: Opracowanie własne.

### 3.4 Korelacja i przyczynowość

W poprzednich podrozdziałach pokazane jest istnienie wahań o częstościach cyklu koniunkturalnego w wielu szeregach dotyczących energii elektrycznej. Niniejszy podrozdział natomiast ma pokazać istnienie zleżności przyczynowo-skutkowej między energią elektryczną a stanem gospodarki. Najlepszym wyznacznikiem zagregowanego stanu gospodarki jest produkt krajowy brutto. W związku z jego kwartalną częstotliwością zaproponowany został jego miesięczny odpowiednik. Miesięczna częstotliwość jest najwyższą, na jaką można bez większych zastrzeżeń rozszacować kwartalne PKB. Wyliczone w podrozdziałach 3.3.1 i 3.3.2 współczynniki korelacji pozwoliły badać istnienie przyczynowości.

Średniookresowe komponenty produkcji i zużycia energii elektrycznej zostały wyodrębnione z szeregów za pomocą filtru Christiano-Fitzgeralda. Wartości korelacji krzyżowej między tymi komponentami a składnikiem cyklicznym miesięcznego PKB jest najwyższa co do modułu dla 3 opóźnień produkcji energii ogółem oraz 2 opóźnień krajowego zużycia energii elektrycznej. Szeregi te wyprzedzają koniunkturę, co jest potwierdzone przez analizę punktów zwrotnych miesięcznych szeregów energii elektrycznej.

Tablica 3.8: Punkty zwrotne szeregów produkcji i zużycia krajowego energii elektrycznej

Szczyt/Dno	Produkcja	Zużycie	MPKB
D	1993.05	-	-
S	1996.02	1996.02	-
D	1998.01	1996.11	1998.12
S	2000.09	1998.11	2000.09
D	2002.02	2002.02	2002.04
S	2004.04	-	2004.04
D	2005.01	-	2005.04
S	2006.01	2006.01	-
D	2007.01	2006.12	-
S	2007.11	2007.11	2008.02
D	2009.04	2009.04	2009.05
S	2011.05	2012.02	2011.04
D	2012.10	-	2013.02

*Źródło:* Opracowanie własne.

Punkty zwrotne produkcji i zużycia energii elektrycznej, zaprezentowane w tablicy 3.8, potwierdzają wyniki korelacji krzyżowej. Dodatkowo, średnia długość cyklu w okresie 2002.01-2013.10 wynosi około 32 miesiące dla produkcji i 40 dla zużycia. Wartości te są zbliżone do długości cykli miesięcznego i kwartalnego PKB - odpowiednio 37 miesiące i 3 lata i  $\frac{1}{4}$  kwartału. Wartości te nie odbiegają bardzo od 2.35 roku uzyskanego z analizy widma danych 15-minutowych.

Według definicji zaproponowanej przez Grangera<sup>60</sup> stacjonarny proces  $x_t$  jest przyczyną

<sup>60</sup>Granger C.W.J., 1969, *Investigating Causal Relations by Econometric Models and Cross-Spectral*

innego procesu stacjonarnego  $y_t$ , gdy  $x_t$  jest lepiej prognozowany z wykorzystaniem  $y_t$  niż bez niego. Zgodnie z definicją przedstawioną w Hamiltonie<sup>61</sup> zmienna  $y$  nie jest przyczyną w sensie Grangera zmiennej  $x$ , jeśli średni błąd prognozy wartości  $x$  nie maleje przy wykorzystaniu opóźnionych wartości zmiennej  $y$ , innymi słowy gdy

$$MSE [\mathbb{E} (x_{t+s} | x_t, x_{t-1}, \dots, y_t, y_{t-1}, \dots)] = MSE [\mathbb{E} (x_{t+s} | x_t, x_{t-1}, \dots)] \quad (3.17)$$

dla wszystkich  $s > 0$ , gdzie MSE oznacza średni błąd prognozy.

Literatura poświęcona badaniu przyczynowości i zależności między energią a wzrostem gospodarczym jest bardzo szeroka. Analizy poświęcone są przypadkom bardzo wielu krajów, np.: Wolde-Rufael<sup>62</sup> dla 17 krajów afrykańskich, Squalli<sup>63</sup> dla 11 krajów eksportujących ropy naftowej czy Lee i innych<sup>64</sup> dla 22 krajów OECD, oraz są zróżnicowane pod względem wykorzystanych metod, np.: przyczynowość Grangera w Kraft i Kraft<sup>65</sup>, badanie kointegracji w Abosedra i Baghestani<sup>66</sup>, wykorzystanie analizy falkowej w Aslan i innych<sup>67</sup>, czy model keynesowski w Narayan i innych<sup>68</sup>. Warto tutaj także przytoczyć artykuł Yu i Choi<sup>69</sup>, w którym w grupie analizowanych krajów znalazła się też Polska.

Obszerna i szczegółowa analiza literatury poświęconej zagadnieniom istnienia oraz kierunku zależności przyczynowo-skutkowej między energią a PKB przedstawiona jest w Ozturk<sup>70</sup>. Choć wiedza o występowaniu związku między tymi zmiennymi może być niezwykle przydatna dla polityków i decydentów, to nie istnieją jednoznaczne wnioski potwierdzające uniwersalny jego kierunek. W zależności od użytego zbioru danych, horyzontu czasowego oraz zastosowanej metody wnioski się różnią. Ozturk<sup>71</sup> przytacza bada-

---

*Methods, Econometrica*, 37(3), s. 424–438.

<sup>61</sup>Hamilton J.D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, str. 303.

<sup>62</sup>Wolde-Rufael Y., 2006, *Electricity Consumption and Economic Growth: A Time Series Experience for 17 African Countries*, *Energy Policy*, 34(10), s. 1106–1114.

<sup>63</sup>Squalli J., 2007, *Electricity Consumption and Economic Growth: Bounds and Causality Analyses for OPEC members*, *Energy Economics*, 29(6), s. 1192–1205.

<sup>64</sup>Lee C.-C., Chang C.-P., Chen P.-F., 2008, *Energy Consumption and GDP in Developing Countries: A Cointegrated Panel Analysis*, *Energy Economics*, 27(3), s. 415–427.

<sup>65</sup>Kraft J., Kraft A., 1978, *On the Relationship Between Energy and GNP*, *Journal of the Energy Development*, 3(2), s. 401–403.

<sup>66</sup>Abosedra S., Baghestani H., 1989, *New Evidence on the Causal Relationship between U.S. Energy, Consumption and Gross National Product*, *Journal of Energy Development*, 14, s. 285–292.

<sup>67</sup>Aslan A., Apergis N., Yildirim S., 2014, *Causality between Energy Consumption and GDP in the U.S.: Evidence from Wavelet Analysis*, *Frontiers in Energy*, 8(1), s. 1–8.

<sup>68</sup>Narayan P.K., Narayan S., Smyth R., 2011, *Energy Consumption at Business Cycle Horizons: The Case of the United States*, *Energy Economics*, 33, s. 161–167.

<sup>69</sup>Yu E.S.H., Choi J.Y., 1985, *The Causal Relationship between Energy and GNP: An International Comparison*, *Journal of Energy and Development*, 10, s. 249–272.

<sup>70</sup>Ozturk I., 2010, *A Literature Survey on Energy-Growth Nexus*, *Energy Policy*, 38(1), s. 340–349.

<sup>71</sup>Ozturk I., 2010, *A Literature Survey...*

nia dla USA, m.in. Krafta i Krafta<sup>72</sup>, Sterna<sup>73</sup>, Lee<sup>74</sup>) oraz Akarcy i Longa<sup>75</sup>, w których uzyskano wnioski odpowiednio o tym, że PKB powoduje energię elektryczną, energia jest przyczyną PKB, o istnieniu wzajemnej przyczynowości i w końcu jej braku. Wniosków nie można nawet pogrupować w żaden określony sposób, na przykład tak, że dla krajów rozwiniętych PKB jest przyczyną energii, a w rozwijających potwierdzana jest hipoteza odwrotna. W badaniu Krafta i Krafta<sup>76</sup>, które zapoczątkowało analizy tej zależności, postawione zostały cztery hipotezy.

1. Hipoteza o neutralności (*neutrality hypothesis*) według której nie zachodzi związek przyczynowo-skutkowy między energią a stanem gospodarki oraz zmienne te są nieskorelowane. W konsekwencji żadna polityka odnośnie sektora energetycznego nie będzie miała odzwierciedlenia we wzroście gospodarczym i odwrotnie.
2. Hipoteza powściągliwości<sup>77</sup> (*conservation hypothesis*) jest potwierdzana, gdy dowodzona jest przyczynowość od wzrostu gospodarczego do energii elektrycznej. W konsekwencji żadna polityka zmniejszania zużycia energii nie będzie wpływać na PKB. Zmiany w zagregowanej produkcji natomiast będą miały swoje odzwierciedlenie w popycie na energię.
3. Hipoteza wzrostu (*growth hypothesis*) nie może być odrzucona, gdy zużycie energii elektrycznej wpływa na PKB.
4. Hipoteza wzajemnej zwrotności (*feedback hypothesis*) jest potwierdzana wtedy, gdy obserwowana jest dwukierunkowa zależność przyczynowo-skutkowa.

Przyczynowość Grangera została przetestowana za pomocą modeli wektorowej autoregresji, oszacowanych w programie Stata. Odsezonowane zmienne zostały zróżnicowane a następnie pozbawione średnich. Oszacowane zostały dwa modele VAR z siedmioma opóźnieniami, a liczba opóźnień była dobrana tak, aby w modelu nie pojawiała się autokorelacja reszt. Zgodnie z hipotezą zerową, zmienna z lewej kolumny tablicy 3.9 nie jest przyczyną w sensie Grangera zmiennych w kolejnych kolumnach.

Z analizy przyczynowości wynika, że dla rozważanej częstotliwości miesięczne PKB pozwala uzyskiwać dokładniejsze prognozy zużycia energii elektrycznej. W związku z tym zastosowanie szeregu reprezentującego cykl koniunkturalny jako zmiennej egzogeniczej w modelu dla danych 15-minutowych powinno, zgodnie z przedstawioną definicją, pozwolić uzyskać dokładniejsze prognozy. Szeregi energii elektrycznej, jak zostało pokazane, są wyprzedzające względem sytuacji gospodarczej, lub innymi słowy, to koniunktura jest

<sup>72</sup>Kraft J., Kraft A., 1978, *On the Relationship Between Energy...*

<sup>73</sup>Stern D.I., 1993, *Energy and Economic Growth in the USA: A Multivariate Approach*, Energy Economics, 15(2), s. 137-150.

<sup>74</sup>Lee C.-C., 2006, *The Causality Relationship between Energy Consumption and GDP in G-11 Countries Revisited*, Energy Policy, 34(9), s. 1086-1093.

<sup>75</sup>Akarca A.T., Long T.V., 1980, *On the Relationship between Energy and GNP: A Re-Examination*, Journal of Energy Development, 5, s. 326-331.

<sup>76</sup>Kraft J., Kraft A., 1978, *On the Relationship Between Energy...*

<sup>77</sup>Nie istnieje polskie tłumaczenie dla tej hipotezy, więc jest to moja propozycja w świetle treści hipotezy.

Tablica 3.9: Wyniki testu przyczynowości Grangera

Przyczyna	$\Delta$ MPKB	$\Delta$ Produkcja Energii	$\Delta$ Zużycie Energii
$\Delta$ MPKB	– –	8,3266 (0,305)	20,334 (0,005)
$\Delta$ Produkcja Energii	9,0794 (0,247)	–	–
$\Delta$ Zużycie Energii	6,0839 (0,530)	–	–

W nawiasach znajdują się wartości p-value testu na bycie przyczyną w sensie Grangera.  
*Źródło:* Opracowanie własne.

opóźniona względem produkcji i zużycia energii elektrycznej, więc jako zmiennej egzogenicznej użyć można zaproponowany wyprzedzający wskaźnik koniunktury, który wyprzedza koniunkturę o około trzy miesiące i jest jednocześnie równoczesny do energii.

### 3.5 Zmienne objaśniające

Jednym ze sposobów wykorzystania informacji o stanie gospodarki do modelowania i prognozowania zapotrzebowania na energię elektryczną jest uwzględnienie jej w modelu szeregu czasowego jako zmiennej objaśniającej (*explanatory variable*). Włączanie zmiennych objaśniających do modeli przestrzeni stanów ma długą tradycję, a klasyczny przykład ich wykorzystania przedstawiony został w Harvey i Durbin<sup>78</sup>. Zaprezentowany tam model liczby ofiar wypadków drogowych w Zjednoczonym Królestwie, oprócz komponentów trendu i sezonowego, obejmował także szereg czasowy ceny paliwa i informacji o wprowadzeniu obowiązku zapinania pasów bezpieczeństwa. Podobną analizę dla danych belgijskich wykonali Hermans, Wets i van den Bossche<sup>79</sup>. W ich opracowaniu, oprócz zmiennych opisujących regulacje prawne, pojawiły się warunki pogodowe i liczba rejestrowanych samochodów. Harvey, Koopman i Penzer<sup>80</sup> posłużyli się szeregami dochodu na osobę i względnych cen napojów alkoholowych jako zmiennymi objaśniającymi w modelowaniu spożycia napojów alkoholowych per capita w Zjednoczonym Królestwie.

Model przestrzeni stanów, w którym zmienna obserwowana jest sumą stochastycznego trendu i trygonometrycznej sezonowości, jest rozszerzeniem modelu (2.69)-( ) o zmienną

<sup>78</sup>Harvey A.C., Durbin J., 1986, *The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling*, Journal of the Royal Statistical Society series B, 149, s. 187–227.

<sup>79</sup>Hermans E., Wets G., Van den Bossche F., 2006, *Frequency and Severity of Belgian Road Traffic Accidents Studied by State-Space Methods*, Journal of Transportation and Statistics, 9(1), s. 63–76.

<sup>80</sup>Harvey A.C., Koopman S.J., Penzer J., *Messy Time Series: A Unified Approach*, Advances in Econometrics, 13, s. 103–143.



objaśniającą  $x_t$  i ma postać

$$y_t = \mu_t + \sum_{i=1}^k \gamma_{i,t} + \beta x_t + \varepsilon_t, \quad (3.18)$$

$$\mu_t = \mu_{t-1} + \nu_{t-1}, \quad (3.19)$$

$$\nu_t = \nu_t + \zeta_t, \quad (3.20)$$

$$\gamma_{i,t} = \gamma_{i,t-1} \cos(\lambda_i) + \gamma_{i,t-1}^* \sin(\lambda_i) + \omega_{i,t}, \quad (3.21)$$

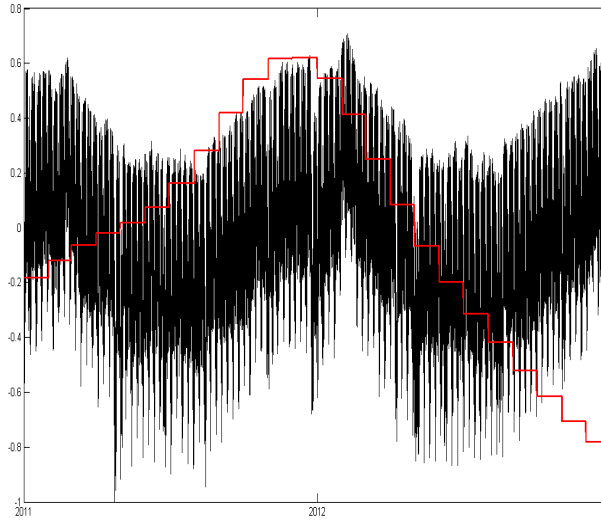
$$\gamma_{i,t}^* = -\gamma_{i,t-1} \sin(\lambda_i) + \gamma_{i,t-1}^* \cos(\lambda_i) + \omega_{i,t}^*, \quad (3.22)$$

gdzie  $\beta$  jest nieznanym współczynnikiem. Estymacja modelu (3.18)-(3.22) jest analogiczna do estymacji modeli przedstawionych w rozdziale 2. Wprowadzenie zmiennej objaśniającej do modelu przekłada się na zmianę w filtrze Kalmana w etapie predykcji, w którym wartość błędu  $v_{t|t-1}$  jest wyznaczana w sposób

$$v_{t|t-1} = y_t - H\alpha_{t|t-1} - A'X_t,$$

gdzie  $X_t$  jest macierzą zmiennych objaśniających.

Rysunek 3.8: Zmienna objaśniająca skonstruowana z komponentu cyklicznego miesięcznego PKB



*Źródło:* Opracowanie własne.

Wyestymowany w rozdziale 2 model potrójnej sezonowości trygonometrycznej z deterministycznym trendem, jak już zostało zauważone, nie bierze pod uwagę wpływu wahań średniokresowych na popyt na energię elektryczną. Uwzględnienie oddziaływania cyklu koniunkturalnego jest możliwe w takim modelu na przykład przez dodanie do niego zmiennej objaśniającej. Wyestymowane w podrozdziale 3.2 miesięczne PKB zostało użyte jako zmienna objaśniająca w modelach wysokoczęstotliwościowych zapotrzebowania na

Tablica 3.10: Wyniki estymacji modeli zawierających miesięczne PKB jako zmienną objaśniającą

	Model z podwójną sezonowością		Model z potrójną sezonowością	
	+ MPKB		+ MPKB	
$10^5 \sigma_\varepsilon^2$	2,692	2,691	1,334	1,333
$10^5 \sigma_\eta^2$	3,636	3,636	0	0
$10^5 \sigma_\omega^2$	3,467	3,473	139,138	118,196
$\beta$		-0,011		-0,012
$10^{-3} \log L$	2,135126	2,135127	2,225767	2,225769
$R^2$	53,141%	53,142%	79,260%	79,263%

Źródło: Opracowanie własne.

energię elektryczną. Miesięczne PKB mają częstotliwość miesięczną i jest to najwyższa częstość z jaką, nie budząc kontrowersji, można aproksymować wahania koniunkturalne. W związku z tym, miesięczne wskaźniki zostały dodane do modelu 15-minutowego w taki sposób, że dla każdego miesiąca jego wartość została powtórzona dla wszystkich kwadransów. Oznacza to, że wartości  $x$  ma postać funkcji schodkowej i jej wartości zmieniają się tylko w momentach zmiany miesiąca. Ilustracją tego jest rysunek 3.8, na którym wyrysowane zostały znormalizowane i pozbawione średnich szeregi logarytmu zapotrzebowania na energię i zmiennej  $x$  dla lat 2011-2012.

Z powodu uwzględnienia trendu w modelowanych danych, zmienne objaśniające zostały uwzględnione w wersji stacjonarnej - jako swoje komponenty cykliczne. Tablica 3.10 prezentuje oszacowania modeli z dodatkową zmienną obrazującą koniunkturę dla okresu 2011-2012. Niewielka zmiana wartości współczynnika determinacji oznacza, że dołączenie do modelu zmiennej objaśniającej nie poprawia znacznie dopasowania modelu do danych. Formalne przetestowanie istotności parametru  $\beta$  wykonane zostało za pomocą testu ilorazu wiarygodności. Wartości statystyk testowych dla modeli z podwójną i potrójną strukturą sezonową wyniosły odpowiednio 0,2 i 0,4, co przy wartości krytycznej równej 3,84 nie daje podstaw<sup>81</sup> do odrzucenia hipotezy o nieistotności parametru  $\beta$ . Prawdopodobną tego przyczyną może być długość próbki, a model ze zmienną aproksymującą wahania średniookresowe należałoby oszacować dla większego zakresu czasowego danych.

<sup>81</sup>Odpowiadające statystykom wartości p-value były równe odpowiednio 0,6547 i 0,5271.

# Rozdział 4

## Zmienne egzogeniczne

W poprzednich rozdziałach wykorzystane zostały jedynie informacje zawarte w szeregu czasowym zapotrzebowania na moce przesyłowe krajowych sieci energetycznych. Oprócz analizy składników cyklu, trendu i sezonowego istotne z punktu widzenia modelowania mogą być informacje zewnętrzne. Najoczywistszym jest uwzględnienie zmiennych pogodowych.

Warunki atmosferyczne mają wpływ zarówno na wiele decyzji podmiotów. O ile część wahań popytu na energię może być przypisana do stałych efektów kalendarzowych, na przykład do godziny wschodu i zachodu słońca, to inna część, przypisywana warunkom atmosferycznym, dla przykładu zachmurzeniu, nie może mieć deterministycznej postaci. Intuicja ekonomiczna wskazuje na istnienie istotnej zależności między wyborami gospodarujących podmiotów a bieżącymi i prognozowanymi warunkami atmosferycznymi. Spędzenie weekendowego popołudnia w parku czy na działce wydaje się bardziej prawdopodobne podczas pięknej, słonecznej, złotej polskiej jesieni niż w chłodny deszczowy dzień końca września. Dodatkowo, przyjemna temperatura wieczorem może skłaniać do zorganizowania ogniska lub grilla, co także wpływa na schemat zużycia energii. Wydaje się, że sobotnie pochmurne popołudnie spędzone w domu powinno się przełożyć na wzrost poboru energii elektrycznej w stosunku do dnia spędzonego poza domem przy pozostałych warunkach niezmiennych. Wysokie temperatury latem i niskie zimą mogą przekładać się na wzrost zapotrzebowania na wzrost popytu na energię ze względu na wykorzystanie klimatyzacji i ogrzewania. Uzasadnioną wydaje się być w związku z tym hipoteza o wpływie czynników atmosferycznych na popyt na energię elektryczną.

W bieżącym rozdziale analizowany jest wpływ zmiennych pogodowych na zapotrzebowanie na moce przesyłowe krajowych sieci elektroenergetycznych za pomocą różnych - standardowych i mniej standardowych - modeli. W podrozdziale 4.1 zaprezentowane są dane pogodowe. Podrozdział 4.2 poświęcony jest analizie danych miesięcznych, co spowodowane jest potrzebą dalszej weryfikacji hipotezy o wpływie cyklu koniunkturalnego na energię przesyłaną w sieci. W ramach tej analizy uwzględniłem także informację o warun-

kach atmosferycznych. Zależności między popytem na energię elektryczną a zmiennymi pogodowymi modelowane są w sposób parametryczny i nieparametryczny odpowiednio w podrozdziałach 4.3 i 4.4. Ostatnia część rozdziału 4 zawiera badanie wpływu dni wolnych od pracy na zapotrzebowanie na moce przesyłowe krajowego systemu energetycznego.

## 4.1 Zmienne pogodowe

Informacje pogodowe są powszechnie wykorzystywane w modelowaniu i prognozowaniu cen i popytu na energię. Zmienne pogodowe dla amerykańskich danych wykorzystywali między innymi Engle i inni<sup>1</sup>, Harvey i Koopman<sup>2</sup>, Weron<sup>3</sup>, w modelowaniu francuskiego zapotrzebowania na energię – Dordonnat i inni<sup>4</sup>, danych hiszpańskich – Valor, Meneu i Caselles<sup>5</sup>. Bessec i Fouquan<sup>6</sup> rozważyli zależność między temperaturą a zużyciem energii elektrycznej w 15 europejskich krajach, Al-Zayer i Al-Ibrahim<sup>7</sup> modelowali wpływ temperatury na popyt na energię elektryczną w Arabii Saudyjskiej, natomiast Henley i Peirson<sup>8</sup> modelowali wpływ pogody na popyt na energię elektryczną w Zjednoczonym Królestwie. Istnienie zależności między temperaturą a zużyciem energii elektrycznej w różnych krajach jest argumentem za wykorzystaniem danych pogodowych w modelowaniu i prognozowaniu zapotrzebowania na energię w Polsce.

Źródłem danych pogodowych była strona <http://www.ogimet.com>, której celem jest dostarczanie danych pogodowych w szybki i profesjonalny sposób<sup>9</sup>. Dane są bezpłatne i ogólnodostępne, ale niestety ich wadą jest sposób prezentacji, w którym serwis może zaprezentować jednocześnie zaledwie 50 obserwacji. Na stronie ogimet.com dostępne są informacje dla 76 stacji meteorologicznych w Polsce. Dostępne dane godzinowe zostały ściągnięte za pomocą specjalnie napisanego makra VBA Excela, dla wszystkich godzin dla lat 2002-2013.

<sup>1</sup>Engle R.F., Granger C.W.J., Rice J., Weiss A., 1986, *Semiparametric Estimates of the Relation between Weather and Electricity*, Journal of the American Statistical Association, 81, s. 310-320.

<sup>2</sup>Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity Demand Using Time-varying Splines*, Journal of the American Statistical Association, 88, s. 1228-1237.

<sup>3</sup>Weron R., 2006, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, John Wiley & Sons.

<sup>4</sup>Dordonnat V., Koopman S.J., Ooms M., Dessertaine A., 2008, *An Hourly Periodic State Space Model for Modeling French National Electricity Load*, International Journal of Forecasting, 24, 588-587.

<sup>5</sup>Valor E., Meneu V., Caselles V., 2001, *Daily Air Temperature and Electricity Load in Spain*, Journal of Applied Meteorology, 40, s. 1413-1421.

<sup>6</sup>Bessec M., Fouquan J., 2008, *The Non-Linear Link between Electricity Consumption and Temperature in Europe: A Threshold Approach*, Energy Economics, 30, s. 2705-2721.

<sup>7</sup>Al-Zayer J., Al-Ibrahim A.A., 1998, *Modeling the Impact of the Temperature on Electricity Consumption in Eastern Province of Saudi Arabia*, Journal of Forecasting, 15, s. 97-106.

<sup>8</sup>Henley A., Peirson J., 1997, *Non-Linearities in Electricity Demand and Temperature*, Oxford Bulletin of Economics and Statistics, 59.

<sup>9</sup><http://www.ogimet.com/index.phtml.en>

Tablica 4.1: Fragment bazy danych pogodowych

DATA	Temperatura (C)		Pkt. rosy (C)	Wilg. Med (%)	Wiatr (km/h)		Ciśn. (Hp)	Opad (mm)	Całk. zachm.	Niskie zachm.	Nast. D-1 (h)	Widocz. Km	Śnieg Opad (cm)	Godz.
	Max	Min			Kier.	Siła	Poryw							
2003-01-02	-4,1	-14,9	-9,8	73,2	S	9,7	58,4	0,4	2,3	2,7	3,8	10,9	7	0
2003-01-02	-3,8	-14,9	-9,6	73,1	S	10,1	59,6	0,4	2,5	2,9	3,8	10,9	7	1
2003-01-02	-3,8	-14,9	-9,3	72,7	S	10,3	59,6	0,4	2,6	3,0	3,8	11,0	7	2
2003-01-02	-3,3	-14,9	-9,0	72,8	S	10,8	58,4	0,4	2,9	3,2	3,8	10,9	7	3
2003-01-02	-2,9	-14,6	-8,6	72,8	S	11,3	58,4	0,4	3,2	3,5	3,8	10,8	7	4
2003-01-02	-2,4	-14,5	-8,2	73,0	S	11,9	59,7	0,4	3,5	3,7	3,8	10,7	7	5
2003-01-02	-1,8	-14,5	-7,8	73,3	S	12,5	61,0	1,9	3,9	4,0	5,8	10,5	6	6
2003-01-02	-1,3	-14,4	-7,3	74,0	S	12,8	56,4	1,9	4,1	4,3	5,8	10,3	6	7
2003-01-02	-0,8	-14,1	-6,9	74,8	S	13,0	57,1	1,9	4,5	4,4	5,8	10,2	6	8
2003-01-02	-0,2	-13,9	-6,4	75,7	S	13,3	56,8	1,9	4,9	4,6	5,8	10,0	6	9
2003-01-02	0,6	-13,4	-5,8	76,7	S	13,5	55,8	2,0	5,1	4,8	5,8	9,8	6	10
2003-01-02	1,1	-13,4	-5,4	78,0	S	13,8	56,3	2,0	5,4	5,0	5,8	9,6	6	11
2003-01-02	1,5	-13,4	-5,0	79,4	S	13,9	56,7	3,1	5,7	5,3	5,8	9,2	6	12
2003-01-02	1,9	-13,3	-4,6	80,6	S	14,0	56,9	3,1	5,9	5,5	5,8	8,9	6	13
2003-01-02	2,1	-13,3	-4,3	81,8	S	14,3	56,9	3,1	6,1	5,8	5,8	8,7	6	14
2003-01-02	2,2	-13,3	-4,0	82,9	S	14,5	56,9	3,1	6,3	6,1	5,8	8,5	6	15
2003-01-02	2,3	-9,4	-3,4	83,9	S	14,7	56,9	3,1	6,5	6,3	5,8	8,3	6	16
2003-01-02	2,4	-9,1	-3,1	84,8	S	14,8	56,9	3,1	6,7	6,4	5,8	8,1	6	17
2003-01-02	2,6	-8,9	-2,5	85,8	S	14,8	56,9	3,8	6,9	6,5	5,8	7,8	6	18
2003-01-02	2,7	-8,9	-2,0	86,8	S	14,8	57,2	3,8	7,1	6,6	5,8	7,6	6	19
2003-01-02	2,8	-8,8	-1,7	87,5	S	14,8	57,2	3,8	7,2	6,6	5,8	7,5	6	20
2003-01-02	2,8	-9,1	-1,5	88,3	S	14,6	57,2	3,8	7,4	6,7	5,8	7,3	6	21
2003-01-02	2,8	-9,0	-1,1	89,0	S	14,3	57,5	3,8	7,5	6,7	5,8	7,2	6	22
2003-01-02	2,9	-8,9	-0,8	89,6	S	13,6	56,9	3,8	7,6	6,7	5,8	7,1	6	23
2003-01-03	2,9	-8,8	-0,5	89,8	S	13,1	56,1	4,6	7,7	6,8	5,8	7,0	6	0
2003-01-03	3,0	-8,6	-0,3	90,3	SSW	12,6	56,1	4,6	7,8	6,8	5,8	6,9	6	1
2003-01-03	3,0	-8,4	-0,1	90,4	SSW	12,4	56,5	4,6	7,8	6,8	5,8	6,9	6	2
2003-01-03	3,2	-8,5	0,1	90,7	SSW	11,8	56,3	4,5	7,8	6,8	5,8	6,9	6	3
2003-01-03	3,3	-3,4	0,6	90,9	SSW	11,3	56,7	4,4	7,8	6,9	5,8	7,0	6	4
2003-01-03	3,4	-3,0	0,8	91,1	SSW	10,9	57,4	4,4	7,8	6,9	5,8	7,0	6	5
2003-01-03	3,4	-2,7	0,9	91,3	SSW	10,2	57,1	5,1	7,8	6,9	0,2	7,1	8	6

Źródło: <http://www.ogimet.com>

Dane przedstawiają sumaryczny obraz stanu pogody dla wszystkich, dostępnych serwisowi ogimet.com stacji meteorologicznych. Zawarte w nich są informacje o temperaturze (maksimum, minimum i mediana), temperaturze punktu rosy, wilgotności, wietrze (kierunku, intensywności i porywie), ciśnieniu atmosferycznym, wielkości opadów, poziomie zachmurzenia, nasłonecznieniu, widoczności poziomej i pokrywie śnieżnej. Fragment danych został przedstawiony w tablicy 4.1.

Tablica 4.2: Podstawowe statystyki zmiennych pogodowych

Zmienna	Liczba obs.	Średnia	Odch. std.	Min	Max
<i>lowclocta</i>	103211	4,34	1,74	0,00	7,70
$ \Delta lowclocta $	103209	0,05	0,09	0,00	3,30
<i>temp_med</i>	103211	8,47	8,31	-20,70	26,50
$ \Delta temp\_med $	103209	0,09	0,09	0,00	4,40
<i>visibility</i>	103211	14,90	6,27	2,40	35,00
$ \Delta visibility $	103209	0,14	0,24	0,00	4,30

Źródło: Opracowanie własne.

W przedstawionych w dalszej części pracy wykorzystałem medianę temperatury, zachmurzenie chmurami piętra niskiego<sup>10</sup>, widoczności i nasłonecznieniu. W modelach dla danych 15-minutowych za warunki atmosferyczne dla poszczególnych kwadransów użyłem wartości zaobserwowanych dla godziny. Jestem świadomy, że jest to pewne przybliżenie, jednak analizy danych pogodowych<sup>11</sup> wykazały małą zmienność wartości w horyzoncie 2 godzin. Argumentem jest także to, że dane są wartościami średnimi lub medianami danych uzyskanych dla 76 stacji meteorologicznych. Dodatkowym uzasadnieniem jest także to, że w polskim klimacie ekstremalne zjawiska pogodowe zdarzają się niezwykle rzadko.

## 4.2 Analiza miesięczna

Konwersja danych wysokoczęstotliwościowych na niskoczęstotliwościowe zazwyczaj polega na zsumowaniu obserwacji z odpowiednich okresów. W przypadku danych o wykorzystaniu mocy przesyłowych krajowych sieci elektroenergetycznych postępowanie musi być odmienne. Zmienna ta nie jest zmienną przepływu lecz stanu i w związku z tym należy postępować jak na przykład z temperaturą czy poziomem wody w rzece - czyli wybrać jedną z obserwacji z agregowanego okresu. Często wykorzystywana w takich przypadkach „reguła kciuka” polega na wyborze obserwacji z trzeciej środy każdego miesiąca. Postępowanie to ma zapewnić lepszą porównywalność danych między miesiącami oraz usunięcie części - niechcianych i zaburzających wnioskowanie - efektów sezonowych. W literaturze

<sup>10</sup>Na podstawie wstępnych analiz wyciągnąłem wniosek o istotności zachmurzenia chmurami piętra niskiego oraz o nieistotności całkowitego zachmurzenia.

<sup>11</sup>Odchylenia standardowe pierwszych i drugich różnic szeregów widoczności, zachmurzenia i temperatury nie przekraczały wartości 0,3.

zabieg ten można najczęściej spotkać w publikacjach dotyczących testowania hipotezy o efektywności rynku finansowego. Bigman, Goldfarb i Schechtman<sup>12</sup> analizowali efektywność rynku, gdzie między innymi cena spot pochodziła z trzeciej środy miesiąca zakończenia kontraktu. Anderson, Duru i Reeb<sup>13</sup> użyli danych z trzeciej środy miesiąca do wyliczenia spreadu akcji w analizie korporacyjnej nieprzejrzystości spółek w Stanach Zjednoczonych. Wybór ten został tam uzasadniony wyeliminowaniem braków danych spowodowanych świętami oraz minimalizacją efektu sezonowego dni tygodnia.<sup>14</sup> Dodatkowego argumentu dostarczają także publikacje Agencji Rynku Energii<sup>15</sup>, w których prezentowane są dobowe wykresy zapotrzebowania na moce przesyłowe KSE w trzecią środę miesiąca.

Jak już zostało wspomniane na początku tej części pracy poświęconej zmiennym egzogenicznym, modele przedstawione w tym podrozdziale są modelami dla danych miesięcznych. Postępowanie to ma pozwolić zweryfikować hipotezę o sensowności wykorzystania informacji o cyklu koniunkturalnym w modelowaniu popytu na energię elektryczną w szerszym zakresie czasowym, ale przy użyciu klasycznego narzędzia – regresji liniowej. Tablica 4.3 przedstawia oszacowania parametrów modeli regresji liniowej dla danych miesięcznych oraz standardowy i skorygowany współczynnik  $R^2$  i  $R_a^2$ . Wiersze White, Ramsey i BGodfrey zawierają wartości p-value dla testowania hipotez odpowiednio o homoskedastyczności reszt, poprawności formy funkcyjnej oraz braku autokorelacji reszt odpowiedniego rzędu.

We wszystkich poniższych przypadkach modelowaną zmienną  $y$  jest logarytm kwadransowego zapotrzebowania na moce przesyłowe KSE w GWh. Oszacowane zostały modele regresji liniowej postaci:

$$y_t = \beta_0 + \beta_1 t + \beta_2 \text{miesiac}_t^2 + \dots + \beta_{12} \text{miesiac}_t^{12} + \beta_{13} \text{coin}_t + \beta_{14} t * \text{coin}_t + \dots + \beta_{15} \text{miesiac}_t^1 * t + \dots + \beta_{26} \text{miesiac}_t^{12} * t + \varepsilon_t, \quad (4.1)$$

gdzie  $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$ , a zmienna *coin* oznacza komponent cykliczny wskaźnika koniunktury przedstawionego w rozdziale 3. Przedstawione modele zawierają komponent deterministycznej sezonowości modelowanej za pomocą zmiennych zero-jedynkowych oraz deterministycznego liniowego trendu ( $t$ ). Zmienna dla stycznia została pominięta, ze względu na problem współliniowości ze stałą i jako taka jest poziomem bazowym. Modele (1) i (2) z tablicy 4.3 miały pozwolić na wnioskowanie o istotności cyklu koniunkturalnego w modelowaniu zapotrzebowania na moce przesyłowe, jednak ze względu na niskie wartości p-value, w testach na poprawność formy funkcyjnej i homoskedastyczność, wnioskowanie na ich podstawie jest nieuprawnione. Rozszerzenie modelu o interakcję między deterministycznym trendem liniowym a zmiennymi zero-jedynkowymi dla poszczególnych miesięcy pozwala rozwiązać problemy, występujące wcześniej, z niepoprawną formą funkcyjną i

<sup>12</sup>Bigman D., Goldfarb D., Schechtman E., 1983, *The Journal of Futures Markets*, 3(3), s. 321-334.

<sup>13</sup>Anderson R.C., Duru A., Reeb D.M., 2009, *Founders, Heirs, and Corporate Opacity in the United States*, *Journal of Financial Economics*, 92, s. 205-222.

<sup>14</sup>Tamże, str. 209.

<sup>15</sup>ARE 2004, *Statystyka elektroenergetyki polskiej*, Agencja Rynku Energii, Warszawa, str. 44.

Tablica 4.3: Miesięczne modele regresji

	(1)	(2)	(3)	(4)	(5)
	y	y	y	y	y
t	0,001***	0,001***	0,001***	0,001***	0,001***
luty	−0,019	−0,019	−0,120	−0,120	−0,118
marzec	−0,074***	−0,074***	−0,321	−0,321	−0,317
kwiecień	−0,137***	−0,138***	−0,327	−0,327	−0,320
maj	−0,167***	−0,168***	−0,586*	−0,585*	−0,576*
czerwiec	−0,179***	−0,179***	−0,698**	−0,697**	−0,686**
lipiec	−0,180***	−0,181***	−0,664**	−0,663**	−0,650**
sierpień	−0,231***	−0,232***	−0,146	−0,145	−0,131
wrzesień	−0,153***	−0,154***	−0,444	−0,443	−0,427
październik	−0,083***	−0,084***	−0,149	−0,147	−0,135
listopad	−0,050**	−0,051**	−0,051	−0,049	−0,033
grudzień	−0,000	−0,001	0,070	0,072	0,094
coin		0,298**	0,475	0,297**	
t*coin			−0,000		
styczeń*t			0,000	0,000	0,000
luty*t			0,000	0,000	0,000
marzec*t			0,000	0,000	0,000
kwiecień*t			0,000	0,000	0,000
maj*t			0,001	0,001	0,001
czerwiec*t			0,001*	0,001*	0,001*
lipiec*t			0,001*	0,001*	0,001*
sierpień*t			−0,000	−0,000	−0,000
wrzesień*t			0,001	0,001	0,000
październik*t			0,000	0,000	0,000
listopad*t			0,000	−0,000	−0,000
grudzień*t			−0,000	−0,000	−0,000
Stała	9,161***	9,196***	9,389***	9,386***	9,341***
$R^2$	0,845	0,855	0,875	0,875	0,865
$R_a^2$	0,830	0,840	0,848	0,849	0,838
N	141,000	141,000	141,000	141,000	141,000
Ramsey	0,003	0,037	0,360	0,304	0,011
White	0,002	0,025	0,906	0,938	0,883
BGodfrey					
L1	0,022	0,114	0,207	0,206	0,038
L2	0,007	0,063	0,099	0,986	0,010
L12	0,323	0,622	0,155	0,153	0,050

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ 

Źródło: Opracowanie własne.



braku stałości wariancji.

Modele (3) i (4) przedstawione w tablicy 4.3, spełniające założenia klasycznego modelu regresji liniowej, umożliwiają wnioskowanie o sensowności używania szeregu reprezentującego koniunkturę do modelowania popytu na energię. Model (3), zawierający iloczyn cyklu i trendu, pozwala wyciągnąć wniosek o nieistotności tej interakcji i stwierdzić niezależność cyklu i trendu. Na podstawie oszacowań modelu (4) stwierdzić należy, że parametr przy zmiennej *coin* jest statystycznie różny od zera, na poziomie istotności 5%, wobec czego zmienna ta jest istotna. Dopełniającego argumentu za potrzebą uwzględniania koniunktury dostarcza model (5), w którym nałożone zostało ograniczenie  $\beta_{coin} = 0$ . Usunięcie zmiennej *coin* prowadzi do problemów autokorelacji i heteroskedastyczności reszt.

Kolejnym etapem analizy danych miesięcznych było uwzględnienie w modelach regresji danych pogodowych. Z teorii ekonometrii wynika, że nawet w przypadku pominięcia w modelu istotnych zmiennych można uzyskać zgodne i nieobciążone oszacowania parametrów, o ile pominięte zmienne są ortogonalne do tych już uwzględnionych<sup>16</sup>. Wydaje się bezsprzecznym, że stan aury jest niezależny od długookresowej ścieżki wzrostu, jak i cyklu koniunkturalnego. Zauważyć należy jednakże, że pogoda, choć ma wpływ na obserwowane wartości produktu krajowego brutto (na przykład ciepła zima wpływa na wzrost w budownictwie, co z kolei przekłada się na PKB), jednak nie może być rozumiana jako czynnik sprawczy zmiany fazy cyklu koniunkturalnego. Modele regresji z informacjami o stanie aury nie są sprzeczne z modelami zaprezentowanymi w tablicy 4.3, a pominięte tam zmienne pogodowe nie zaburzają wnioskowania.

Model (4) z tablicy 4.3 został rozszerzony o dane o medianie temperatury (*temp\_med*), zachmurzeniu (*lowclocta*), widoczności poziomej (*visibility*) oraz liczbie godzin słonecznych (*sun*). W analizie wpływu danych atmosferycznych referencyjnym jest model (4) z tablicy 4.3, a jego oszacowania zostały zaprezentowane w tablicy 4.4 pod postacią modelu (1). Spośród przedstawionych tam modeli (2)-(4), zawierające zmienne pogodowe, na podstawie skorygowanego  $R^2$  należałoby wybrać model (2) ze wszystkimi zmiennymi pogodowymi. Usuwanie nieistotnych, na poziomie istotności 5%, zmiennych prowadzi do obniżenia wartości  $R_a^2$ . W związku z tym w dalszej analizie uwzględnione zostały wszystkie wspomniane zmienne pogodowe.

Bez wątplenia temperatura pozwala lepiej tłumaczyć zmienność zapotrzebowania na moce przesyłowe, natomiast kwestia ważności pozostałych zmiennych będzie analizowana w modelach korzystających z wyższych częstotliwości. Być może, wpływ pozatemperaturowych zmiennych pogodowych okaże się istotny w przypadku danych o wyższej częstotliwości. Wyniki powyższych modeli nie przesądzają oczywiście o istotności informacji o stanie koniunktury i pogody dla danych wysokoczęstotliwościowych. Dają jednak podstawę do dalszego wykorzystania zmiennych reprezentujących cykl koniunkturalny i informacji o warunkach meteorologicznych w dalszym modelowaniu.

<sup>16</sup>Greene W.H., 2003, *Econometric Analysis 5ed*, Prentice Hall, rozdział 8.2.1 str. 148.

Tablica 4.4: Miesięczne modele regresji z danymi pogodowymi

	(1)	(2)	(3)	(4)
	y	y	y	y
t	0,001***	0,001***	0,001***	0,001***
coin	0,297**	0,304**	0,321***	0,320***
temp_med		-0,003**	-0,003**	-0,004**
lowcloudocta		0,006	0,005	
visibility		0,002		
sun		0,003*	0,004*	0,003
luty	-0,120	-0,090	-0,072	-0,029
marzec	-0,321	-0,257	-0,282	-0,253
kwiecień	-0,327	-0,239	-0,293	-0,276
maj	-0,585*	-0,522*	-0,540*	-0,525*
czerwiec	-0,697**	-0,583**	-0,593**	-0,591**
lipiec	-0,663**	-0,591**	-0,615**	-0,594**
sierpień	-0,145	-0,010	-0,062	-0,087
wrzesień	-0,443	-0,417	-0,415	-0,386
październik	-0,147	-0,046	-0,048	-0,055
listopad	-0,049	-0,013	-0,006	-0,007
grudzień	0,072	0,075	0,063	0,103
luty*t	0,000	0,000	0,000	0,000
marzec*t	0,000	0,000	0,000	0,000
kwiecień*t	0,000	0,000	0,000	0,000
maj*t	0,001	0,001	0,001	0,001
czerwiec*t	0,001*	0,001*	0,001*	0,001*
lipiec*t	0,001*	0,001*	0,001*	0,001*
sierpień*t	-0,000	-0,000	-0,000	-0,000
wrzesień*t	0,001	0,001	0,001	0,000
październik*t	0,000	-0,000	-0,000	0,000
listopad*t	-0,000	-0,000	-0,000	-0,000
grudzień*t	-0,000	-0,000	-0,000	-0,000
Stała	9,386***	9,328***	9,357***	9,373***
$R^2$	0,875	0,890	0,888	0,885
$R_a^2$	0,849	0,863	0,861	0,859
N	141,000	141,000	141,000	141,000

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ 

Źródło: Opracowanie własne.

## 4.3 Analiza danych dziennych

Rozważania z poprzedniego rozdziału dają przesłanki do uwzględnienia informacji o cyklu koniunkturalnym i temperaturze w modelowaniu zapotrzebowania na moce przesyłowe sieci energetycznych. Zwiększenie częstotliwości, z miesięcznej na dzienną, ma na celu łatwiejsze wychwycenie występujących w danych zależności oraz w mniejszym stopniu argumentowane jest szybkością estymacji modeli z mniejszą liczbą obserwacji. Podstawą analizy było 4290 obserwacji z godziny 11:15 dla wszystkich dni kalendarzowych ze zbioru danych.

Tablica 4.5 zawiera oszacowania trzech modeli regresji dla danych dziennych. Pierwszy z nich jest próbą jak najprostszego modelowania tych obserwacji, ale niestety hipoteza o poprawności formy funkcyjnej jest silnie odrzucana (p-value testu Ramsey'a równe 0). Intuicyjnie fakt ten można zinterpretować w ten sposób, że wzrost częstotliwości danych powinien przekładać na wzrost skomplikowania modelu. Model (2) jest odpowiednikiem modelu (2) z tablicy 4.4 estymowanym dla dziennych danych, jednak z analogicznego powodu co model (1) jest on odrzucany. Dołączenie interakcji między dniem a miesiącem, ze względu na odmienny wpływ dni w poszczególnych miesiącach, pozwoliło uzyskać model (3) o poprawnej formie funkcyjnej<sup>17</sup>.

Warte podkreślenia jest potwierdzenie się przypuszczenia z podrozdziału 4.2 o tym, że pozatemperaturowe zmienne pogodowe mogą okazać się istotne dla danych o wyższej częstotliwości. Zrozumiałe są znaki oszacowań przy poziomie zachmurzenia (*lowclouds*) i nasłonecznieniem (*sun*). Intuicyjne okazało się ponadto oszacowanie przy widoczności (*visibility*), które należy interpretować jako spadek zapotrzebowania na moce przesyłowe KSE średnio o 0,1% wraz z jednostkowym wzrostem poziomu widoczności.

Zwraca uwagę, zarówno w tablicy 4.4 jak i 4.5, znak oszacowania parametru przy medianie temperatury. W modelu (4) wartość oszacowania -0,005 oznacza, że wzrost temperatury o jeden stopień Celsjusza przełoży się średnio na spadek zapotrzebowania na moce przesyłowe o 0,5%. Wniosek wydaje się być uzasadniony podczas jesieni, zimy i wiosny, jednak latem sytuacja może być odmienna. Gdy tylko temperatura przewyższa pewien próg, zaobserwować można wzrost poboru energii elektrycznej związanej najprawdopodobniej z działaniem urządzeń klimatyzacyjnych. W związku z tym, uzasadnione wydaje się być odwołanie do bardziej zaawansowanych technik modelowania nieliniowego wpływu temperatury na popyt na energię elektryczną.

---

<sup>17</sup>W modelu tym spełnione jest też założenie o braku autokorelacji reszt.

Tablica 4.5: Modele regresji dla danych dziennych

	(1) y	(2) y	(3) y
t	0,000***	0,000***	0,000***
coin	0,484***	0,494***	0,494***
temp_med	-0,004***	-0,004***	-0,005***
lowclocta	0,004***	0,003***	0,003***
sun	0,002***	0,002***	0,002***
visibility	-0,000	-0,001***	-0,001***
poniedziałek	0,236***	0,236***	0,248***
wtorek	0,254***	0,254***	0,244***
środa	0,260***	0,260***	0,257***
czwartek	0,251***	0,251***	0,244***
piątek	0,250***	0,250***	0,239***
sobota	0,157***	0,157***	0,144***
luty	0,002	-0,014	-0,015
marzec	-0,032***	-0,049***	-0,055***
kwiecień	-0,084***	-0,110***	-0,125***
maj	-0,111***	-0,160***	-0,157***
czerwiec	-0,087***	-0,150***	-0,157***
lipiec	-0,072***	-0,136***	-0,152***
sierpień	-0,081***	-0,139***	-0,148***
wrzesień	-0,065***	-0,111***	-0,120***
październik	-0,026***	-0,048***	-0,057***
listopad	-0,023***	-0,035***	-0,033*
grudzień	-0,021***	-0,015	0,004
luty*t		0,000*	0,000*
marzec*t		0,000*	0,000*
kwiecień*t		0,000***	0,000***
maj*t		0,000***	0,000***
czerwiec*t		0,000***	0,000***
lipiec*t		0,000***	0,000***
sierpień*t		0,000***	0,000***
wrzesień*t		0,000***	0,000***
październik*t		0,000**	0,000**
listopad*t		0,000	0,000
grudzień*t		-0,000	-0,000
poniedziałek & luty			-0,007
⋮			⋮
sobota & grudzień			-0,011
Stała	9,586***	9,627***	9,631***
$R^2$	0,805	0,815	0,821
$R_a^2$	0,804	0,814	0,816
N	4290,000	4290,000	4290,000
Ramsey	0,000	0,003	0,050

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ 

Źródło: Opracowanie własne.

## 4.4 Nieliniowe zależności

Rozgraniczenie w literaturze przedmiotu na stopniodni grzania (*heating degree day*) i stopniodni chłodzenia (*cooling degree day*) świadczy o nieliniowym przełożeniu temperatury na zapotrzebowanie na energię. Nieliniowa zależność między tymi zmiennymi w literaturze pojawia się najczęściej w postaci V-kształtnej. Engle i inni<sup>18</sup>, Harvey i Koopman<sup>19</sup>, Valor, Meneu i Caselles<sup>20</sup> oraz Dordonnat i inni<sup>21</sup> uzyskiwali dla danych ze Stanów Zjednoczonych, Hiszpanii i Francji taką właśnie zależność.

Na tej podstawie można wysunąć hipotezę o tym, że także w przypadku danych polskich taką V-kształtność zależności powinno się zaobserwować. Wzrost zapotrzebowania na energię elektryczną w okresie letnim związany jest z wysokimi temperaturami, które powodują wzrost wykorzystania urządzeń chłodzących, w tym klimatyzacji, przez co wzrasta zapotrzebowanie na energię. Argumentem przeciwnym jest niska liczba domów i mieszkań wyposażonych w urządzenia klimatyzacyjne<sup>22</sup>. Według statystyki Głównego Urzędu Statystycznego urządzenia klimatyzacyjne znajdowały się w 0,35% gospodarstw domowych<sup>23</sup>, podczas gdy w dane amerykańskie wskazują na analogiczny odsetek wynoszący prawie 83%<sup>24</sup>. Podejrzewać można jednak, że w biurach, urzędach i centrach handlowych w Polsce wykorzystuje się dużą liczbę urządzeń klimatyzacyjnych, więc w dalszej części rozważam hipotezę o nieliniowości relacji między temperaturą a zapotrzebowaniem na energię.

Nieliniowa zależność może być analizowana w modelu semiparametrycznym lub poprzez liniowe przybliżenie w modelu regresji. Zaprezentuję najpierw wyniki dwóch modeli semiparametrycznych, po czym zaproponuję sposób modelowania zależności między popytem na energię elektryczną a temperaturą w liniowym modelu szacowanym metodą najmniejszych kwadratów. Model regresji semiparametrycznej ma postać

$$y_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_K x_{K,i} + f(z_i) + \varepsilon_i, \quad (4.2)$$

dla  $i = 1, 2, \dots, N$ , gdzie  $y$  jest zmienną objaśnianą,  $x_1, \dots, x_K$  są zmiennymi niezależnymi,  $\varepsilon$  jest zaburzeniem z zerową średnią i stałą wariancją  $\sigma_\varepsilon^2$ , natomiast  $z$  jest zmienną

<sup>18</sup>Engle R.F., Granger C.W.J., Rice J., Weiss A., 1986, *Semiparametric Estimates...*

<sup>19</sup>Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity Demand...*

<sup>20</sup>Valor E., Meneu V., Caselles V., 2001, *Daily Air Temperature...*

<sup>21</sup>Dordonnat V., Koopman S.J., Ooms M., Dessertaine A., 2008, *An Hourly Periodic State Space Model...*

<sup>22</sup>Przesłanką potwierdzającą hipotezę mógłaby być statystyka urządzeń klimatyzacyjnych w biurach, urzędach i lokalach handlowo-usługowych. Dane dla Stanów Zjednoczonych publikowane są przez Energy Information Administration (U.S. Energy Information Administration, 2009, *Commercial Building Energy Consumption Survey*), natomiast ich odpowiednik dla Polski nie istnieje.

<sup>23</sup>GUS, 2014, *Zużycie energii w gospodarstwach domowych w 2012 r.*, Główny Urząd Statystyczny, Warszawa, str. 36.

<sup>24</sup>U.S. Energy Information Administration, 2009, *Residential Energy Consumption Survey*, tablica HC 7.1 *Air Conditioning in U.S. Homes, by Housing Unit Type*: Na 113,6 mln gospodarstw domowych dane wskazują, że 94 mln posiadało sprawne i używało urządzenia klimatyzacji.

objaśniającą wpływającą na analizowaną zmienną  $y$  w sposób nieliniowy, który może być opisany funkcją  $f$  dowolnej postaci. Po wyestymowaniu parametrów  $\beta$  można dopasować funkcję  $f$  w równaniu

$$y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \dots - \hat{\beta}_K x_{K,i} = f(z_i) + \varepsilon_i, \quad (4.3)$$

dla  $i = 1, 2, \dots, N$ . Po oszacowaniu parametrów  $\beta$  model (4.3) staje się modelem nieparametrycznym i wiele różnych technik może zostać użytych do jego estymacji. W najogólniejszym przypadku estymowaną zależność  $f$  będącą warunkową wartością oczekiwaną można zapisać

$$f(z) = \mathbb{E}(Y|Z = z) = \int y \frac{f(z, y)}{f_Z(z)} dz = \frac{\int y f(z, y) dz}{f_Z(z)} \quad (4.4)$$

Dla zbioru obserwacji  $\{Z_i, Y_i\}$  dla  $i = 1, \dots, N$  wykorzystując estymator jądrowy  $\hat{f}_{h,g}(\cdot)$

$$\hat{f}_{h,g}(x, z) = \frac{1}{N} \sum_{i=1}^N K_h \quad (4.5)$$

uzyskuje się

$$\int y \hat{f}_{h,g}(z, y) dy = \frac{1}{N} \sum_{i=1}^N K_h(z - Z_i) Y_i \quad (4.6)$$

gdzie  $K_h, K_g$  są ustalonymi jądrami. Estymatorem  $f_Z(z)$  jest

$$\hat{f}_h(z) = \frac{1}{N} \sum_{i=1}^N K_h(z - Z_i) \quad (4.7)$$

gdzie  $K_h$  jest ustalonym jądrem. Odwołując się do 4.6 i 4.7 estymator  $\hat{f}(z)$  można zapisać jako

$$\hat{f}_h = \frac{n^{-1} \sum_{i=1}^N K_h(z - Z_i) Y_i}{n^{-1} \sum_{j=1}^N K_h(z - Z_j)}. \quad (4.8)$$

Estymacja parametrów  $\beta$  w równaniu 4.2 może być dokonana za pomocą różnych technik. Poniżej przedstawione estymacje zostały wykonane za pomocą dwóch z nich – Robinsona<sup>25</sup> i Yatchewa<sup>26</sup> – ze względu na większą odporność wyników. Wykorzystane zostały implementacje w środowisku Stata zaproponowane przez odpowiednio Verardiego i Debarskiego<sup>27</sup> oraz Lokshina<sup>28</sup>. Szczegółowy opis modeli nieparametrycznych i semiparamet-

<sup>25</sup>Robinson P.M., 1998, *Root-N-Consistent Semiparametric Regression*, *Econometrica* 66, s. 931-954.

<sup>26</sup>Yatchew A., 1997, *An Elementary Estimator of the Partial Linear Model*, *Economic Letters* 57, s. 135-143.

<sup>27</sup>Verardi V., Debarsy N., 2012, *Robinson's Square Root-of-N-Consistent Semiparametric Regression Estimator in Stata*, *Stata Journal*, 12(4), s. 726-735.

<sup>28</sup>Lokshin M., 2006, *Difference-Based Semiparametric Estimation in Partial Linear Models*, *Stata Journal*, Vol. 6(3), s. 377-383.

trycznych znajduje się w Hardle i inni<sup>29</sup>.

Estymator podwójnych reszt (*double residual estimator*) Robinsona<sup>30</sup> umożliwia oszacowanie modelu 4.2 poprzez zastosowanie różnicy między wartością obserwowaną  $y$  a jej warunkową wartością oczekiwaną, to jest

$$\begin{aligned} y_i - \mathbb{E}(y_i|z_i) &= \beta_0 + \beta_1 x_{1,i} + \dots + f(z_i) + \varepsilon_i - (\beta_0 + \beta_1 \mathbb{E}(x_{1,i}|z_i) + \dots + f(z_i)) \\ &= \beta_1 [x_{1,i} - \mathbb{E}(x_{1,i}|z_i)] + \dots + \beta_K [x_{K,i} - \mathbb{E}(x_{K,i}|z_i)] + \varepsilon_i \end{aligned} \quad (4.9)$$

dla  $i = 1, 2, \dots, N$ . Jeśli znane są warunkowe wartości oczekiwane  $\mathbb{E}(y_i|z_i)$  i  $\mathbb{E}(x_{h,i}|z_i)$ , dla  $h = 1, 2, \dots, K$ , to parametry równania (4.9) mogą zostać oszacowane za pomocą metody najmniejszych kwadratów. W przeciwnym przypadku są one szacowane metodą najmniejszych kwadratów z równania postaci

$$y_i - \hat{m}_y(z_i) = \beta_1 (x_{1,i} - \hat{m}_{x_1}(z_i)) + \dots + \beta_K (x_{K,i} - \hat{m}_{x_K}(z_i)) + \varepsilon_i \quad (4.10)$$

gdzie  $\hat{m}_y(z_i)$  i  $\hat{m}_{x_{h,i}}(z_i)$  są zgodnymi estymatorami  $y_i = \hat{m}_y(z_i) + \varepsilon_{1,i}$  oraz  $x_i = \hat{m}_{x_{h,i}}(z_i) + \varepsilon_{2h,i}$ . Estymatorem wektora parametrów  $\beta$  w równaniu (4.9), zgodnie z klasycznym wzorem na estymator w metodzie najmniejszych kwadratów, jest

$$\hat{\beta} = \left( (\mathbb{X} - \hat{M}_{\mathbb{X}}(z))' (\mathbb{X} - \hat{M}_{\mathbb{X}}(z)) \right)^{-1} (\mathbb{X} - \hat{M}_{\mathbb{X}}(z))' (Y - \hat{M}_{\mathbb{Y}}(z)). \quad (4.11)$$

W sposobie estymacji równania (4.2) zaproponowanym przez Yatchewa<sup>31</sup> zakłada się o funkcji  $f$ , że jest ciągła, jednowartościowa, a jej pochodna jest skończona. Zaproponowany przez Yatchewa sposób estymacji parametru  $\beta$  polega na posortowaniu obserwacji w porządku rosnącym i estymacji modelu (4.2) na pierwszych różnicach zamiast na poziomach, to jest

$$y_n - y_{n-1} = (f(z_n) - f(z_{n-1})) + \beta_1 (x_{1,n} - x_{1,n-1}) + \dots + \beta_K (x_{K,n} - x_{K,n-1}) + \varepsilon_n - \varepsilon_{n-1} \quad (4.12)$$

dla  $n = 2, 3, \dots, N$ . Wraz ze wzrostem liczebności próby,  $f(z_n) - f(z_{n-1}) \rightarrow 0$ , ponieważ pochodna  $f$  jest skończona. Ze względu na występowanie w równaniu (4.12) różnicy błędów losowych, estymator parametrów  $\beta$  nie jest efektywny. Efektywność może być poprawiona poprzez użycie różnic wyższego rzędu, a uogólnienie równania (4.12) przedstawione jest w Yatchew<sup>32</sup> lub Lokshin<sup>33</sup>.

Tablica 4.6 zawiera oszacowania modeli z nieliniową zależnością między temperaturą a popytem na energię. Zaprezentowane tam modele (1) i (2) zostały wyestymowane z wy-

<sup>29</sup>Hardle W.K., Muller M., Sperlich S., Werwatz A, 2004, *Nonparametric and Semiparametric Models, An Introduction*, Springer.

<sup>30</sup>Robinson P.M., 1998, *Root-N-Consistent...*

<sup>31</sup>Yatchew A., 1997, *An Elementary Estimator...*

<sup>32</sup>jw.

<sup>33</sup>Lokshin M., 2006, *Difference-Based...*

korzystaniem procedur odpowiednio Robinsona i Yatchewa, natomiast (3) i (4) modelowane w sposób parametryczny przy użyciu metody najmniejszych kwadratów. Model (3) jest modelem regresji z funkcją łamaną, a (4) można obrazowo ująć jako „kij hokejowy”<sup>34</sup> lub małe „l” „pisane”. Niech  $p_t$  będzie sumą wszystkich zmiennych pozatemperaturowych wpływających na zapotrzebowanie na moce przesyłowe

$$p_t = \beta_0 + \beta_1 t + \beta_2 \text{coin}_t + \beta_3 \text{dzien}_t + \beta_4 \text{miesiac}_t + \beta_5 \text{dzien}_t * \text{miesiac}_t + \\ + \beta_6 t * \text{miesiac}_t + \beta_7 \text{lowcloct}_t + \beta_8 \text{sun}_t + \beta_9 \text{visibility}_t, \quad (4.13)$$

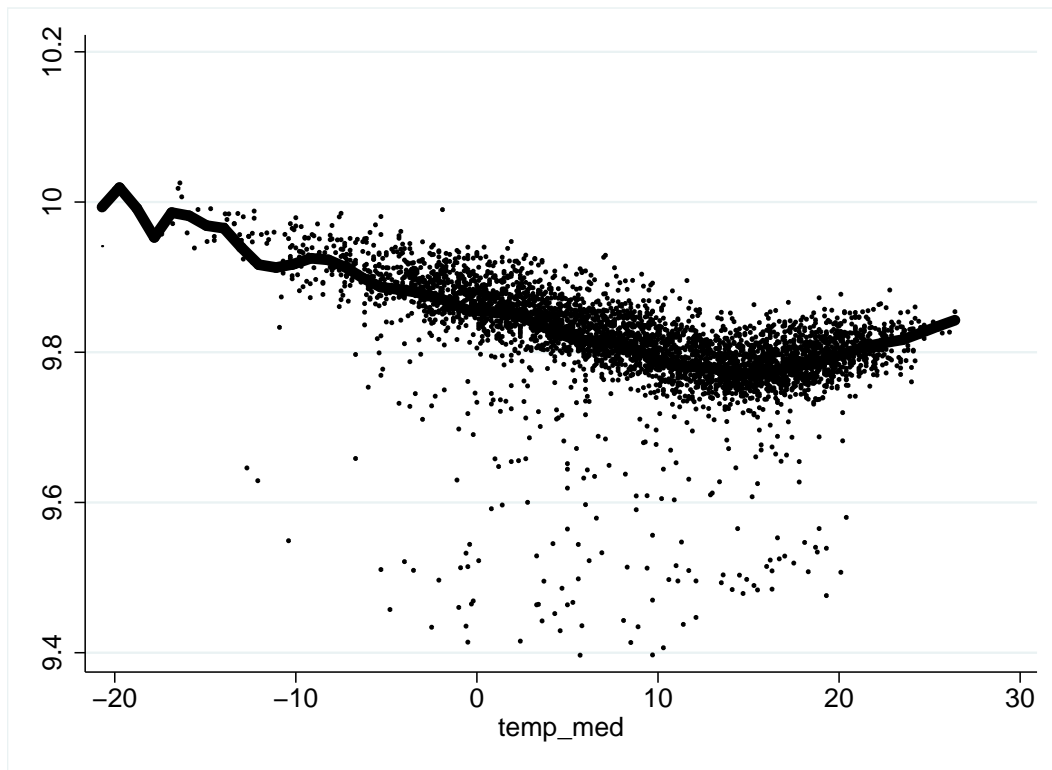
wówczas model (3) można zapisać jako

$$y_t = p_t + \beta_{10} \text{temp\_med}_t + \beta_{11} \mathbb{1}_{(\text{temp\_med}_t > 16)} + \beta_{12} \mathbb{1}_{(\text{temp\_med}_t > 16)} * \text{temp\_med}_t + \varepsilon_t, \quad (4.14)$$

a model (4) w postaci

$$y_t = p_t + \beta_{10} \text{temp\_med}_t + \mathbb{1}_{(14 < \text{temp\_med}_t < 18)} (\beta_{11} + \beta_{12} (\text{temp\_med}_t - 16)^2) + \\ + \mathbb{1}_{(\text{temp\_med}_t > 18)} (\beta_{13} + \beta_{14} \text{temp\_med}_t) + \varepsilon_t. \quad (4.15)$$

Rysunek 4.1: Nieliniowy wpływ temperatury dla danych dziennych



Źródło: Opracowanie własne.

<sup>34</sup>Weron R., 2006, *Modeling and Forecasting...*, str. 74 określił w ten sposób zależność między temperaturą a zapotrzebowaniem na energię elektryczną.



Tablica 4.6: Dzielne modele z nielinową zależnością między energią a temperaturą

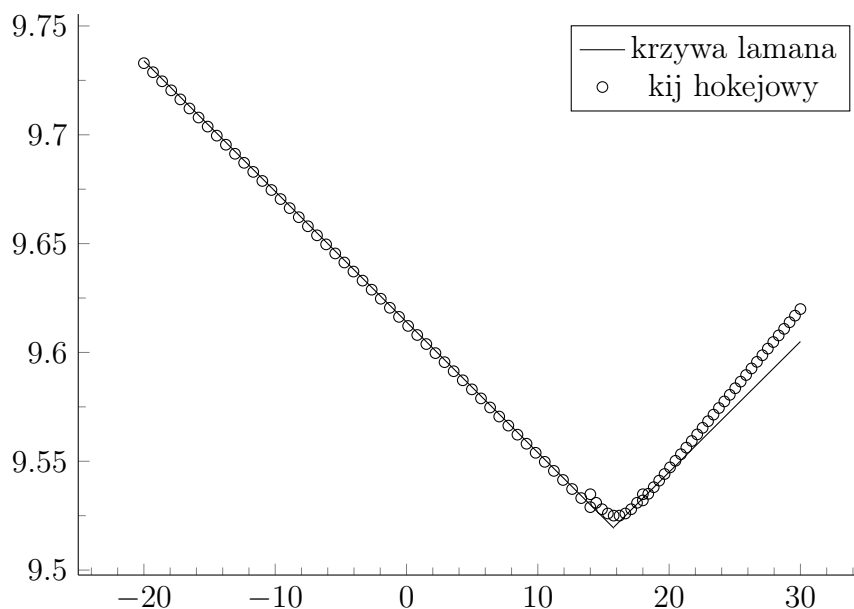
	(1) y	(2) y	(3) y	(4) y
t	0,000***	0,000***	0,000***	0,000***
coin	0,516***	0,433***	0,509***	0,511***
lowclocta	0,006***	0,005***	0,005***	0,005***
sun	0,001*	0,001	0,001*	0,001*
visibility	-0,000	-0,001***	-0,000	-0,000
poniedziałek	0,245***	0,251***	0,248***	0,249***
wtorek	0,239***	0,238***	0,246***	0,246***
środa	0,256***	0,266***	0,260***	0,260***
czwartek	0,241***	0,234***	0,246***	0,246***
piątek	0,235***	0,236***	0,239***	0,239***
sobota	0,142***	0,141***	0,144***	0,144***
luty	-0,012	-0,022	-0,011	-0,010
marzec	-0,047***	-0,048**	-0,044**	-0,043**
⋮	⋮	⋮	⋮	⋮
grudzień	0,004	0,004	0,008	0,009
poniedziałek & luty	-0,006	0,007	-0,007	-0,008
poniedziałek & marzec	-0,010	-0,010	-0,014	-0,014
poniedziałek & kwiecień	-0,052**	-0,058**	-0,055***	-0,056***
⋮	⋮	⋮	⋮	⋮
sobota & grudzień	-0,007	0,016	-0,010	-0,010
luty*t	0,000	0,000	0,000	0,000
marzec*t	0,000**	0,000*	0,000*	0,000*
kwiecień*t	0,000***	0,000**	0,000***	0,000***
maj*t	0,000***	0,000***	0,000***	0,000***
czerwiec*t	0,000***	0,000***	0,000***	0,000***
lipiec*t	0,000***	0,000***	0,000***	0,000***
sierpień*t	0,000***	0,000***	0,000***	0,000***
wrzesień*t	0,000***	0,000***	0,000***	0,000***
październik*t	0,000***	0,000*	0,000***	0,000***
listopad*t	0,000*	0,000	0,000*	0,000*
grudzień*t	-0,000	-0,000	-0,000	-0,000
temp_med			-0,006***	
$\mathbb{1}_{(temp\_med_t > 16)}$			-0,189***	
$\mathbb{1}_{(temp\_med_t > 16)}temp\_med$			0,012***	
$\mathbb{1}_{(temp\_med_t < 14)}temp\_med$				-0,006***
$\mathbb{1}_{(14 \leq temp\_med_t \leq 18)}$				-0,088***
$\mathbb{1}_{(14 \leq temp\_med_t \leq 18)}(temp\_med-16)^2$				0,002
$\mathbb{1}_{(temp\_med_t > 18)}$				-0,213***
$\mathbb{1}_{(temp\_med_t > 18)}temp\_med$				0,007***
Stała			9,614***	9,613***
$R^2$	0,785	0,752	0,831	0,831
$R_a^2$	0,780	0,746	0,827	0,827
N	4290,000	4289,000	4290,000	4290,000

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Zaprezentowane w tablicy 4.6 modele, będące rozszerzeniem (4) z tablicy 4.5, potwierdzają hipotezę o istotności cyklu koniunkturalnego oraz nieliniowym wpływie temperatury w przypadku danych dziennych<sup>35</sup>. Nieliniowość ta jest zaprezentowana na rysunku 4.1, który prezentuje wynik estymacji modelu (4.2) metodą Robinsona. Pokazany został tylko wykres uzyskany z metody Robinsona ze względu na niemalże całkowite jego podobieństwo do relacji wyestymowanej metodą Yatchewa. Oszacowania modelu (3) i (4) potwierdzają hipotezę o nieliniowości zależności między temperaturą a popytem na energię, a ich wizualizacje zamieszczone są na rysunku 4.2.

Potwierdzone zostało ponadto przypuszczenie o V-kształtności relacji, ponieważ dla chłodnych dni spadek temperatury o 1 stopień Celsjusza powoduje średnio spadek zapotrzebowania na energię o 0,6%, natomiast w ciepłe dni oczekujemy, że wzrostowi temperatury o 1°C towarzyszyć będzie wzrost zapotrzebowania na energię o 0,6%. Bardzo podobne wnioski mogą być wyciągnięte na podstawie oszacowań modelu (4), w którym charakter zależności zmienia się bardziej gładko, to jest w otoczeniu temperatury 16°C zgodnie z funkcją kwadratową. Obserwowany jest analogiczny wzrost podczas dni grzewczych, podczas gdy dla temperatur przekraczających 18°C wzrost temperatury prowadzi do wyższego wzrostu zapotrzebowania na moce przesyłowe, to jest o 0,7% przy 0,6% wcześniej. Zauważyć należy, że na podstawie oszacowań uzyskanych dla próby danych dziennych nie można wnioskować, który ze sposobów parametrycznego uwzględniania temperatury jest lepszy, dlatego, że obydwa modele mają równe wartości współczynników determinacji.

Rysunek 4.2: Wykres oszacowanych parametrycznych nieliniowych zależności

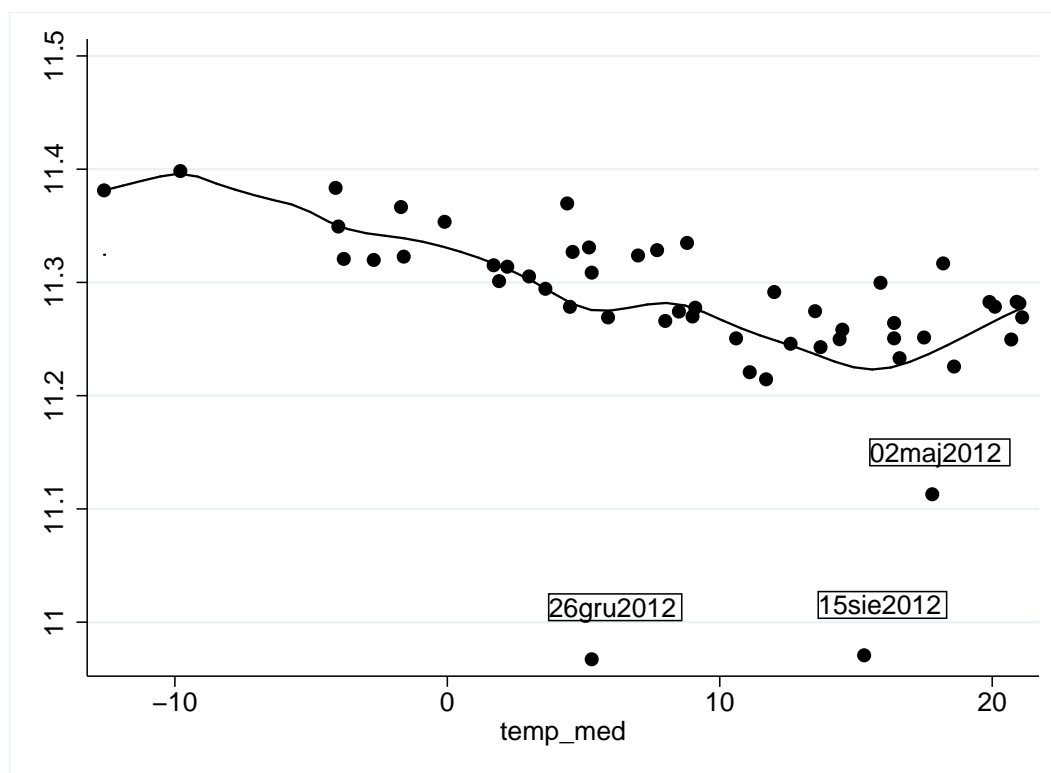


Źródło: Opracowanie własne.

<sup>35</sup>Nie jest to widoczne bezpośrednio w tablicy 4.6 na podstawie oszacowań, ponieważ relacja była estymowana nieparametrycznie.

Intrygujące na rysunku 4.1 wydają się być obserwacje z istotnie mniejszymi wartościami zapotrzebowania przy różnych poziomach temperatury. Dla lepszego zobrazowania tej kwestii sporządzony został rysunek 4.3 prezentujący nieliniową zależność temperatury i potrzebnych mocy przesyłowych krajowych sieci energetycznych dla danych tygodniowych 2012 roku. Analiza obserwacji odstających pozwala powiązać je z dniami wolnymi od pracy (15 sierpnia i 26 grudnia) oraz długimi weekendami (2 maja). Wydaje się, że pełniejszy obraz zapotrzebowania na energię elektryczną można uzyskać przy wykorzystaniu informacji o dniach wolnych.

Rysunek 4.3: Nieliniowy wpływ temperatury dla danych tygodniowych z 2012 roku



Źródło: Opracowanie własne.

## 4.5 Wpływ dni wolnych od pracy

W Polsce dni wolne od pracy określone są ustawowo i związane są ze świętami kościelnymi (Święto Trzech Króli, Wielkanoc, Zielone Świątki, Boże Ciało, Wniebowzięcie Najświętszej Maryi Panny, Uroczystość Wszystkich Świętych i Boże Narodzenie), państwowymi (Święto Państwowe 1 Maja, Święto Narodowe Trzeciego Maja, Narodowe Święto Niepodległości) oraz dodatkowo Nowym Rokiem. Na zużycie energii elektrycznej wpływ mają także długie weekendy, czyli te weekendy, dla których dzień wolny od pracy przypada we wtorek lub czwartek, analogicznie do Dordonnat i innych<sup>36</sup>. Obserwowane wówczas jest mniejsze wykorzystanie mocy produkcyjnych związanych z wydłużaniem czasu wolnego poprzez wzięcie jednego dnia urlopu. Wpływ dni wolnych od pracy na kategorie ekonomiczne został zauważony przez statystyków i ekonometryków, a jedną z tego konsekwencji było pojawienie się odsezonowanych względem liczby dni roboczych szeregów czasowych. Tablica 4.7 zawiera daty analizowanych dalej dni wolnych od pracy.

W literaturze poświęconej szacowaniu popytu na energię elektryczną szeroko wykorzystywane są zmienne mające modelować oddziaływanie dni wolnych od pracy na zapotrzebowanie na moce przesyłowe sieci. Dordonnat i inni<sup>37</sup>, oprócz uwzględniania świąt i długich weekendów, wprowadzili zmienną  $W_t^{28}$  określającą czas letni. Niestandardowe dni uwzględnione w analizie pojawiły się także między innymi u Harvey'a i Koopmana<sup>38</sup>, Henley'a i Peirsona<sup>39</sup>, Pardo, Meneu i Valora<sup>40</sup> oraz Cancelo, Espasa i Grafe<sup>41</sup>.

Tablica 4.8 zawiera oszacowania modelu 4.14 dla danych dziennych w wersjach ze zmiennymi dyskretnymi dla dni ustawowo wolnych od pracy. Wprowadzone zmienne dyskretne przyjmują wartość 1 tylko w przypadku, gdy święto lub część długiego weekendu wypada w dzień roboczy. Model (2) obejmujący wszystkie zmienne dla dni wolnych charakteryzuje się bardzo wysoką wartością współczynnika determinacji  $R^2$ . Niestety, w tym przypadku wyniki testu Ramsey'a skłaniają do odrzucenia hipotezy o poprawności formy funkcyjnej. Poszukiwania modelu spełniającego założenie klasycznego modelu regresji liniowej o właściwej formie funkcyjnej wiążą się w tym przypadku z odrzucaniem zmiennych dyskretnych dla dni wolnych. Modele (3)-(5) są niektórymi z uzyskanych udoskonaleń. Powyższe postępowanie nie świadczy bynajmniej o nielogiczności wykorzystania zmiennych o dniach wolnych od pracy. Przeciwnie, na podstawie modeli (4) i (5) można stwierdzić, że użycie tych informacji pozwala uzyskać lepsze dopasowanie pod względem kryterium  $R_a^2$ <sup>42</sup>.

<sup>36</sup>Dordonnat V., Koopman S.J., Ooms M., Dessertaine A., 2008, *An Hourly Periodic State Space Model...*

<sup>37</sup>Dordonnat V., Koopman S.J., Ooms M., Dessertaine A., 2008, *An Hourly Periodic State Space Model...*

<sup>38</sup>Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity Demand...*

<sup>39</sup>Henley A., Peirson J., 1997, *Non-Linearities in Electricity Demand...*

<sup>40</sup>Pardo A., Meneu V., Valor E., 2002, *Temperature and Seasonality Influences on Spanish Electricity Load*, Energy Economics, 24, s. 55-70.

<sup>41</sup>Cancelo J.R., Espasa A., Grafe R., 2008, *Forecasting the Electricity Load from One Day to One Week Ahead for Short-Term Forecasting*, International Journal of Forecasting, 24, s. 603-615.

<sup>42</sup>Poprawę zauważyć można także na podstawie wartości  $p$  testu Reset Ramsey'a

Zastosowanie zmiennych dla pierwszego i szóstego stycznia, pierwszego i trzeciego maja, jedenastego listopada i długich weekendów bez wątpienia jest kluczowe w modelowaniu dziennego zapotrzebowania na moce przesyłowe KSE. Wykorzystanie pozostałych informacji o dniach wolnych ustawowo zostanie sprawdzone także w przypadku danych o wyższej częstotliwości.

Wyniki uzyskane z modeli parametrycznych i semiparametrycznych wskazują na nieliniowy charakter zależności między temperaturą a zapotrzebowaniem na energię elektryczną. W związku z tym pierwsza główna hipoteza badawcza mojej pracy musi zostać zweryfikowana pozytywnie.

Na podstawie oszacowań modeli z tablicy 4.8 stwierdzam, że pod względem skorygowanego  $R^2$  wybrać należałoby model (2) zawierający zmienne zero-jedynkowe dla wszystkich dni ustawowo wolnych od pracy. Przy przyjęciu za kryterium wyboru modelu spełnienia przez model wymogu poprawności formy funkcyjnej wybrać należałoby model (4) lub (5). Wybór najlepszego modelu do prognozowania wartości zapotrzebowania na moce przesyłowe krajowego systemu elektroenergetycznego spośród (3)-(5) nie jest oczywisty. Z jednej strony modele (4) i (5) spełniają założenie o poprawności formy funkcyjnej, z drugiej strony model (2) charakteryzuje się wyższym współczynnikiem  $R^2$  dzięki czemu może pozwalać wyznaczać dokładniejsze prognozy.

Tablica 4.7: Dni wolne od pracy i długie weekendy w latach 2002-2014

Rok	Nowy Rok	Trzech Króli	Wielkanoc	Majówka	Zielone Świątki	Boże Ciało	Wniebowz. NMP	Wszyst. Świąt.	Święto Niepodl.	Boże Narodzenie	Długie Weekendy
2002	1 I	-	31 III 1 IV	1 3 V	19 V	30 V	15 VIII	1 XI	11 XI	25-26 XII	2V 31V 16VIII
2003	1 I	-	20-21 IV	1 3 V	8 VI	19 VI	15 VIII	1 XI	11 XI	25-26 XII	2V 20VI 10XI
2004	1 I	-	11-12 IV	1 3 V	30 V	19 VI	15 VIII	1 XI	11 XI	25-26 XII	2II 12XI
2005	1 I	-	27-28 III	1 3 V	15 V	26 V	15 VIII	1 XI	11 XI	25-26 XII	2V 27V 31X
2006	1 I	-	16-17 IV	1 3 V	4 V	15 V	15 VIII	1 XI	11 XI	25-26 XII	2V 5V 14VIII
2007	1 I	-	8-9 IV	1 3 V	27 V	7 VI	15 VIII	1 XI	11 XI	25-26 XII	30IV-4V 8VI 2XI 31XII
2008	1 I	-	23-24 III	1 3 V	11 V	22 V	15 VIII	1 XI	11 XI	25-26 XII	2V 23V 10XI
2009	1 I	-	12-13 IV	1 3 V	31 V	11 VI	15 VIII	1 XI	11 XI	25-26 XII	2I 12VI
2010	1 I	-	4-5 IV	1 3 V	23 V	3 VI	15 VIII	1 XI	11 XI	25-26 XII	12XI
2011	1 I	6 I	24-25 IV	1 3 V	12 V	23 V	15 VIII	1 XI	11 XI	25-26 XII	7I 2V 13V 31X
2012	1 I	6 I	8-9 IV	1 3 V	27 V	7 VI	15 VIII	1 XI	11 XI	25-26 XII	30IV-4V 8VI 2XI 31XII
2013	1 I	6 I	31 III 1 IV	1 3 V	19 V	30 V	15 VIII	1 XI	11 XI	25-26 XII	29IV-2V 31V 16VIII 27XII
2014	1 I	6 I	20-21 IV	1 3 V	8 VI	19 VI	15 VIII	1 XI	11 XI	25-26 XII	2V 20VI 10XI

Źródło: Opracowanie własne.

Tablica 4.8: Dienne modele z nieliniową zależnością między energią a temperaturą i czasem wolnym

	(1) y	(2) y	(3) y	(4) y	(5) y
t	0,000***	0,000***	0,000***	0,000***	0,000***
coin	0,509***	0,516***	0,512***	0,508***	0,510***
temp_med	-0,006***	-0,006***	-0,006***	-0,006***	-0,006***
lowclocta	0,005***	0,004***	0,004***	0,005***	0,005***
sun	0,001*	0,001***	0,001*	0,001*	0,001**
visibility	-0,000	-0,000	-0,000	-0,000	-0,000
poniedziałek	0,248***	0,257***	0,256***	0,257***	0,257***
wtorek	0,246***	0,263***	0,262***	0,262***	0,263***
środa	0,260***	0,268***	0,268***	0,268***	0,268***
czwartek	0,246***	0,268***	0,267***	0,267***	0,267***
piątek	0,239***	0,260***	0,259***	0,259***	0,259***
sobota	0,144***	0,144***	0,144***	0,144***	0,144***
luty	-0,011	-0,008	-0,008	-0,007	-0,007
marzec	-0,044**	-0,017*	-0,042***	-0,041***	-0,041***
kwiecień	-0,103***	-0,057***	-0,104***	-0,101***	-0,102***
maj	-0,129***	-0,127***	-0,130***	-0,127***	-0,127***
czerwiec	-0,136***	-0,134***	-0,136***	-0,134***	-0,133***
lipiec	-0,154***	-0,149***	-0,152***	-0,150***	-0,150***
sierpień	-0,141***	-0,136***	-0,140***	-0,138***	-0,137***
wrzesień	-0,094***	-0,092***	-0,095***	-0,092***	-0,092***
październik	-0,037**	-0,034***	-0,036***	-0,034**	-0,034**
listopad	-0,022	-0,017*	-0,018	-0,017	-0,017
grudzień	0,008	0,013	0,010	0,011	0,011
⋮	⋮	⋮	⋮	⋮	⋮
grudzień*t	-0,000	-0,000*	-0,000	-0,000	-0,000
poniedziałek & luty	-0,007	-0,015	-0,015	-0,016	-0,016
poniedziałek & marzec	-0,014	-0,033***	-0,022	-0,022	-0,022
poniedziałek & kwiecień	-0,055***	-0,047***	-0,058***	-0,058***	-0,058***
⋮	⋮	⋮	⋮	⋮	⋮
sobota & grudzień	-0,010	-0,010	-0,010	-0,010	-0,010
$\mathbb{1}_{(temp\_med_t > 16)}$	-0,189***	-0,169***	-0,166***	-0,171***	-0,180***
$\mathbb{1}_{(temp\_med_t > 16)}temp\_med$	0,012***	0,011***	0,011***	0,011***	0,012***
styczeń 1		-0,444***	-0,444***	-0,443***	-0,443***
styczeń 6		-0,268***	-0,268***	-0,267***	-0,267***
maj 1		-0,229***	-0,229***	-0,229***	-0,222***
maj 3		-0,290***	-0,289***	-0,289***	-0,282***
sierpień 15		-0,281***	-0,281***		
listopad 1		-0,401***	-0,401***		
listopad 11		-0,234***	-0,234***	-0,220***	-0,220***
grudzień 25		-0,392***			
grudzień 26		-0,381***			
Wielkanoc		-0,294***			
Zielone Świątki		-0,000			
Boże Ciało		-0,243***	-0,244***	-0,244***	
Długi weekend		-0,114***	-0,108***	-0,106***	-0,104***
Stała	9,614***	9,615***	9,618***	9,615***	9,613***
$R^2$	0,831	0,944	0,899	0,876	0,869
$R_a^2$	0,827	0,943	0,896	0,873	0,865
N	4290,000	4290,000	4290,000	4290,000	4290,000
Ramsey	0,021	0,000	0,012	0,031	0,033

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ 

Źródło: Opracowanie własne.





## Rozdział 5

# Prognozowanie popytu na energię elektryczną

Przedstawione we wcześniejszych rozdziałach analizy i sposoby modelowania szeregu zapotrzebowania na moce przesyłowe krajowych sieci elektroenergetycznych wskazują na potrzebę uwzględniania wszystkich składowych tego szeregu czasowego. Przedstawione tam analizy pozwalają wyciągnąć wnioski odnośnie sposobu jego modelowania. Potrzeba włączenia do modelu komponentu trendu została pokazana w rozdziale 1, konieczność modelowania potrójnej struktury sezonowej - w rozdziale 2, natomiast wpływ cyklu koniunkturalnego na szeregi energii elektrycznej - w rozdziale 3. Na podstawie rozdziału 4 znany jest wpływ pogodowych zmiennych objaśniających. Wydaje się, że wiedza o źródłach wahań analizowanego szeregu pozwala dokładniej przewidywać jego przyszłe wartości.

Zgodnie z metodologią ekonometrii istnieje rozdźwięk między jak najlepszym, w znaczeniu jak najlepiej dopasowanym, modelem a jak najoszczędniejszym podejściem. Konstruowany model powinien być jednocześnie jak najprostszy, czyli zawierać jak najmniejszą liczbę zmiennych, jak i w jak najwyższym stopniu tłumaczyć wahania analizowanej zmiennej zależnej. Widoczne jest to dobrze na przykładzie tabeli (4.7). Model z najwyższym współczynnikiem determinacji liniowej nie jest optymalny pod względem kryterium skorygowanego  $R^2$ . Świadczy to o tym, że w modelu, który w najwyższym stopniu wyjaśnia zmienność popytu na energię elektryczną, przynajmniej jedna zmienna nie jest istotna. Na podstawie literatury poświęconej modelowaniu i prognozowaniu popytu na energię elektryczną wyższy priorytet przyznać można stopniowi dopasowania mierzonemu współczynnikiem determinacji  $R^2$ . Weron<sup>1</sup> stwierdził, że ze względu na straty finansowe związane z przeszacowaniem lub niedoszacowaniem zapotrzebowania na energię niezwykle znaczenie mają nawet próby zredukowania błędu prognozy choćby o ułamek procenta.

Ze względu na charakter niniejszej pracy, rozważone zostać muszą także kwestie obciąż-

---

<sup>1</sup>Weron R., *Modelling and ...*, str. X

żenia estymatorów i ich ewentualny wpływ na prognozy. Z teorii ekonometrii wynika, że spełnienie założenia o poprawności formy funkcyjnej jest jednoznaczne z nieobciążonością i zgodnością estymatorów, natomiast pozytywne zweryfikowanie hipotez o braku autokorelacji i heteroskedastyczności pozwala na wnioskowanie statystyczne. Poprawność modelu, rozumiana jako pozytywne zweryfikowanie wszystkich wymaganych postulatów, uzasadnia prognozy uzyskiwane z modelu, jednak w określonych przypadkach preferowane może być użycie modelu niespełniającego wszystkich założeń teoretycznych. Pojawia się tutaj kolejny problem wyboru między modelem z pozytywnie zweryfikowanymi własnościami a modelem umożliwiającym szybsze lub ekonomiczniejsze prognozy. Powstaje czysto teoretyczne pytanie o to, czy zawsze model, z którego można uzyskać lepsze prognozy, będzie modelem spełniającym wszystkie założenia. Czy przy prognozowaniu popytu na energię elektryczną można rozluźnić wymóg pozytywnej diagnostyki modelu, jeśli efektem tego byłyby dokładniejsze prognozy? Choć pierwszym wyborem ekonometryka powinna być poprawność statystyczna i ekonometryczna, to podać można przykłady praktycznych rozwiązań kompromisowych. Dla przykładu, w statystyce małych obszarów<sup>2</sup> dopuszczenie obciążoności estymatorów przekłada się na uzyskiwanie oszacowań ze znacząco niższą wariancją. W związku z powyższym w tym rozdziale wykorzystane zostaną także metody naiwne, które będą punktem odniesienia do przedstawionych modeli.

Ostatnim, choć nie najmniej ważnym, wątkiem jest kwestia modelu używanego do modelowania i prognozowania. Pierwszym wyborem zawsze jest model regresji liniowej szacowany metodą najmniejszych kwadratów. Sposób ten umożliwia modelowanie deterministycznych zależności, co nie powinno być zawsze postrzegane jako wada. Umiejętność opisanie danego zjawiska za pomocą komponentów deterministycznych posiada tę zaletę, że większą część zmienności będziemy w stanie przewidzieć. Wydaje się, że optymalnym rozwiązaniem byłoby posiadanie modelu regresji liniowej spełniającego wszystkie założenia i posiadającego wysoki współczynnik determinacji liniowej. Niestety, obserwowane zjawiska ekonomiczne charakteryzują się nie deterministycznym, a stochastycznym charakterem. Analiza komponentu sezonowego 15-minutowego szeregu zapotrzebowania na energię elektryczną w Polsce, przedstawiona w rozdziale 2, potwierdza stochastyczny charakter tego komponentu, natomiast na podstawie rozdziału pierwszego stwierdzić należy obecność stochastycznego trendu. Wydaje się, że argumenty o stochastycznym charakterze trendu i sezonowości wystarczają do stwierdzenia niemożliwości znalezienia poprawnego modelu regresji liniowej dla 15-minutowych danych

Dwie tury zawodów prognostycznych przeprowadzonych w latach 1990-1992 opisane przez

---

<sup>2</sup>Statystyka małych obszarów (small area estimation) jest dziedziną statystyki zajmującą się szacowaniem parametrów dla małych podpopulacji. Mały obszar jest definiowany jako część populacji, która jest w niskim stopniu reprezentowana w próbie. Używanie standardowych narzędzi statystycznych do wyznaczenia charakterystyk małego obszaru wiąże się z dużą wariancją oszacowań. Szczegóły na temat statystyki małych obszarów znaleźć można w Rao N.K.J., 2003, *Small Area Estimation*, John Wiley & Sons, s. 344. W Polsce głównym ośrodkiem zajmującym się statystyką małych obszarów jest Urząd Statystyczny w Poznaniu i Katedra Statystyki Akademii Ekonomicznej w Poznaniu.

Ramanathana i innych<sup>3</sup> stanowiły dla mnie podstawę wyboru modeli do prognozowania. Gospodarzem zawodów była spółka energetyczna Puget Sound Power and Light. Udział w zawodach wzięło 11 zespołów badaczy praktyków i teoretyków prognozowania zapotrzebowania na energię elektryczną. Prognozy były wyliczane o 8:00 dla wszystkich godzin następnego dnia roboczego, natomiast w piątek prognozowano godzinowe zapotrzebowanie na energię elektryczną dla 3 kolejnych dni. Jak zauważyli Ramanathan i inni<sup>4</sup> specyfikacje niektórych modeli nie były upublicznione, natomiast na podstawie wyników zawodów wynika, że zaproponowany przez nich model (EGRV) był bardzo konkurencyjny dla prognoz dla dni roboczych. Model zmiennych w czasie okresowych splajnów kubicznych zaproponowany przez Harvey'a i Koopmana<sup>5</sup> w zauważalnej liczbie przypadków okazał się najlepszy. Dla dni weekendowych bardzo dobrze spisywał się model spółki Puget. W aneksie zamieściłem tablicę ze szczegółowymi wynikami zawodów prognostycznych. Ze względu na brak informacji o części modeli w wyborze najlepszego modelu dla polskich dancych wykorzystałem różne wersje modeli EGRV, ARIMA, modelu okresowych splajnów kubicznych. Proponowanie modelu prognostycznego opiera się także na porównaniu jego rezultatów z prognozami naiwnymi.

W dalszej części tego rozdziału przedstawię szczegóły eksperymentu, na podstawie którego dokonałem wyboru najlepszego modelu. Podrozdziały 5.1-5.5 zawierają specyfikacje modeli, a prognozy z nich uzyskane porównałem w podrozdziale 5.6. Ostatni podrozdział 5.7 zawiera eksperyment za pomocą którego zweryfikowałem drugą główną hipotezę badawczą o przydatności informacji o cyklu koniunkturalnym do prognozowania średniookresowego.

## 5.1 Projekt eksperymentu

Przeprowadzenie eksperymentu porównującego wymaga ustalenia horyzontu prognoz. Literatura przedmiotu, na przykład Weron<sup>6</sup>, dzieli prognozy zapotrzebowania na energię elektryczną na krótko-, średnio- i długo-okresowe. Weron<sup>7</sup> zauważył także, że przypisywanie konkretnej długości horyzontu prognozy w literaturze nie jest jednoznaczne. Zdecydowałem, że w eksperymencie zostaną wykorzystane prognozy na 1, 2 i 3 dni jako prognozy krótkookresowe, oraz prognozy na 3 miesiące służące do porównania własności przewidywań modeli w średnim okresie. Analizy i prognozy długookresowe nie zostały przeanalizowane, ponieważ środek ciężkości analizy musiałby się przesunąć z danych o wysokiej częstotliwości w stronę danych o niższych częstościach, prawdopodobnie kwar-

---

<sup>3</sup>Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-run Forecast of Electricity Loads and Peaks*, International Journal of Forecasting, 13, s. 161-174.

<sup>4</sup>Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-run Forecast...*, str. 168.

<sup>5</sup>Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity Demand Using Time-Varying Splines*, Journal of the American Statistical Association, 88(424), s. 1228-1236.

<sup>6</sup>Weron R., *Modelling and ...*, str. 67

<sup>7</sup>Weron R., *Modelling and ...*, str. 67

talnych i rocznych.

Wydaje się, że najlepszym kryterium do porównania modeli prognostycznych jest sprawdzenie trafności uzyskiwanych z nich prognoz. Przeprowadziłem w związku z tym eksperyment mający na celu porównanie zaobserwowanych wartości zapotrzebowania na moce przesyłowe krajowych sieci elektroenergetycznych z prognozami porównywanych modeli. Porównanie prognoz krótkookresowych polegało na oszacowaniu modelu na próbie od stycznia 2002 do 10 października 2013 w pierwszym kroku i wyznaczeniu prognoz na trzy kolejne dni. W każdym kolejnym kroku okres, na którym szacowany był model, zwiększany był o kolejny dzień. W ten sposób dla każdego modelu przeprowadzonych zostało 445 iteracji, aż do 30 grudnia 2014.

Jak już zostało wspomniane, do porównania modeli wybrane zostały: metoda dnia podobnego, modele regresji, modele ARIMA i model okresowych splajnów kubicznych. Ex ante nie wiadome było, która możliwa specyfikacja każdego modelu będzie najlepsza. Z tego powodu, dla każdej wymienionej klasy modeli wyznaczone zostały prognozy dla kilku wersji modelu.

Literatura zawiera opracowania przedstawiające różne podejścia do dni specjalnych w modelowaniu popytu na energię elektryczną. Weron<sup>8</sup> omówił sposoby korygowania outlierów oraz uzupełniania braków danych. Harvey i Koopman w aplikacji swojej metody posłużyli się danymi ze skorygowanymi obserwacjami dla dni specjalnych<sup>9</sup>. Dni świąteczne i święta państwowe zostały przez nich zastąpione średnią z odpowiednich dni tygodnia poprzedniego i następnego. Ramanathan i inni<sup>10</sup> nie opisali żadnych przekształceń swoich danych.

Zdecydowałem oszacować każdy z modeli w eksperymencie był szacowany dla trzech wariantów bazy danych. Pierwsza z nich zawierała oryginalne i w żaden sposób nieskorygowane dane. W drugim zbiorze danych obserwacje odstające zostały zastąpione uzyskanymi dla nich wartościami dopasowanymi. Ze względu na to, że analizowane dane są bardzo mocno zagregowane w przekroju terytorialnym - łączne zapotrzebowanie dla całego kraju - analiza graficzna nie wskazała potencjalnie odstających obserwacji. W przypadku Polski wyłączenia większych obszarów i pozbawienie większej liczby odbiorców dostępu do elektryczności związane jest zwykle z silnymi wiatrami. Za potencjalnie mające największy wpływ i widoczne być może widoczne w skali makro są okresy kilku dni następujących po przejściu orkanu. W okresie 2002-2014 w Polsce odnotowano 11 przypadków tych silnych wiatrów, a informacje o terminie uderzenia orkanów przedstawione są w tablicy 5.1. Za okres potencjalnie widocznego wpływu orkanu na popyt i podaż energii uznane zostały arbitralnie 3 pełne dni po zjawisku. W trzecim typie zbioru danych dni specjalne, związane ze świętami kościelnymi i państwowymi, zostały zastąpione wartościami dopasowanymi z

<sup>8</sup>Weron R., 2006, *Modeling and Forecasting...*, str. 69.

<sup>9</sup>Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity...*, str. 1232.

<sup>10</sup>Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-run Forecast...*

modelu. Dni ustawowo wolnych od pracy zostały wypisane w tablicy 4.7.

Tablica 5.1: Daty przejścia orkanów w Polsce w latach 2002-2014

L.p.	Nazwa	Data
1	Orkan Pia	18 listopada 2004
2	Orkan Britta	1-2 listopada 2006
3	Orkan Kyrill	18-19 stycznia 2007
4	Orkan Paula	26-27 stycznia 2008
5	Orkan Zizi	22-23 lutego 2008
6	Orkan Emma	1-2 marca 2008
7	Orkan Xynthia	28 lutego-1 marca 2010
8	Orkan Yoda	27-28 listopada 2011
9	Orkan Joachim	16-17 grudnia 2011
10	Orkan Andrea	4-5 stycznia 2012
11	Orkan Ksawery	5-7 grudnia 2013

Źródło: Opracowanie własne na podstawie Lorenc H. i inni (red.), 2012, *Klęski żywiołowe a bezpieczeństwo wewnętrzne kraju*, Instytut Meteorologii i Gospodarki Wodnej, Państwowy Instytut Badawczy, Warszawa, oraz uzupełnione o informacje ze strony [http://pl.wikipedia.org/wiki/Orkany\\_w\\_Polsce](http://pl.wikipedia.org/wiki/Orkany_w_Polsce).

Głównym argumentem za takim skonstruowaniem eksperymentu jest zainteresowanie podmiotów z rynku energii minimalizacją błędów prognoz. Porównywanie prognoz za pomocą matematycznego kryterium średniego względnego błędu prognozy wydaje się być najbardziej obiektywne.

Analizy przedstawione we wcześniejszych rozdziałach wykonane zostały na danych od stycznia 2002 do 10 października 2013. W związku z tym, że nowe dane są dostępne z każdym dniem, a właściwie z każdym kwadrantem, wybrać należało granicę dla dodawania nowych obserwacji do zbioru. Za końcową obserwację uznano wartości zmiennych dla ostatniego kwadranta 2014 roku. Powstałą bazę danych można w dosyć naturalny sposób podzielić na okres próby (*in-sample*) od 1 stycznia 2002 do 10 października 2013 i poza próbą (*out-of-sample*) od 11 października 2013 do 31 grudnia 2014 roku. Uzasadnieniem tego podziału danych jest możliwość testowania prognoz dla danych spoza próby zawierających cały rok kalendarzowy.

Zdefiniowanie eksperymentu wymaga przyjęcia jeszcze dodatkowych założeń. Z oszacowanych na danych z próby (*in-sample*) modeli wyznaczane były prognozy. Jednak z powodu obecności w nich niedeterministycznych zmiennych egzogenicznych opisujących warunki atmosferyczne pojawia się kwestia traktowania ich wartości poza próbą (*out-of-sample*), gdyż w momencie szacowania modelu są one nieznane. Dla tego eksperymentu możliwe są przynajmniej dwa rozwiązania. Pierwsze może polegać na użyciu historycznych prognoz pogody, czyli tych przewidywań, które były dostępne dla okresu ostatniej obserwacji z próby. W eksperymencie zamiast prognoz wykorzystałem zaobserwowanych

wartości zmiennych pogodowych. Zaobserwowane warunki atmosferyczne dla okresu poza próbą są zostały wykorzystane zamiast prognoz<sup>11</sup>. Zdecydowałem się, zgodnie z Ramanathanem i innymi<sup>12</sup>, w prognozowaniu krótkookresowym nie wykorzystywać informacji o cyklu koniunkturalnym, ponieważ w krótkim, 3-dniowym, horyzoncie prognoz sytuacja gospodarcza nie zmienia się.

## 5.2 Metoda dnia podobnego

Metoda dnia podobnego w ogólności polega na wyznaczeniu wartości prognozowanej równej zaobserwowanej, historycznej wartości w dniu o zbliżonych charakterystykach. Literatura przedmiotu wskazuje, że metoda dnia podobnego jest wykorzystywana głównie w modelach sieci neuronowych. Weron<sup>13</sup> zauważył, że zaskakująco użyteczną wersją tej metody jest metoda polegająca na wybraniu opóźnionej o tydzień obserwacji. Procedura taka ma kilka zasadniczych i oczywistych wad. Po pierwsze, nie uwzględnia wpływu zmiennych pogodowych, a po drugie, przypisywanie dniom ustawowo wolnym od pracy wartości z odpowiedniego dnia tygodnia poprzedniego powodować może duże błędy prognoz. Alternatywą pozostaje przeszukiwanie obserwacji historycznych i wybór najbardziej podobnego, pod ustalonymi kryteriami, dnia, dla którego sporządzana jest prognoza. Dodatkową kwestią wartą rozważenia jest kwestia długości okresu przeszukiwania. Ze względu na to, że niemożliwym jest stwierdzenie z całą pewnością, która z wersji metody dnia podobnego jest najlepsza, rozważonych zostało kilka jej wersji.

Pierwszą wersją, zgodnie z przytoczonym przez Weroną<sup>14</sup> argumentem, była najprostsza, czyli metoda naiwna, polegająca na wyborze wartości opóźnionej o tydzień względem okresu prognozy. Drugim analizowanym wariantem było wybieranie najbardziej podobnego kwadransa w okresie sześciomiesięcznym. Najbardziej porównywalny do analizowanego kwadrans musiał przypadać na ten sam rodzaj dnia (roboczy, sobota, niedziela, dzień ustawowo wolny) oraz posiadać podobne charakterystyki pogodowe. Opcja trzecia różniła się od drugiej tylko okresem przeszukiwania - zamiast sześciu miesięcy wybierana była zawsze cała historia obserwacji.

Średnie względne błędy prognozy dla trzech wersji metody dnia podobnego zostały zawarte w tablicy 5.2. Przedstawione wartości należy czytać w taki sposób, że dla prognozy

<sup>11</sup>Taylor J.W. i Buizza R., 2003, *Using Weather Ensemble Predictions in Electricity Demand Forecasting*, International Journal of Forecasting, 19, s. 57-70. pokazali przewagę prognoz popytu na energię elektryczną z wykorzystaniem prognoz wiązką modeli (*ensemble forecast*) nad tradycyjnymi, punktowymi prognozami. Interesujący wydaje się być także pomysł na pokonywanie prognoz wspomniany przez Ramanthana i inni, 1993, *Short-run Forecast of Electricity Loads and Peaks*, International Journal of Forecasting, 13, str. 164. Zagadnienie znaczenia i odmiany prognoz warunków atmosferycznych wydaje się być bardzo szerokie i zasługuje na osobne badanie.

<sup>12</sup>Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-run Forecast...*, str. 164.

<sup>13</sup>Weron R., 2006, *Modelling and ...*, str. 79

<sup>14</sup>j.w.

na 1 okres, w tym przypadku tydzień, metoda naiwna pozwala wyznaczyć wartość przewidywaną, która w stosunku do wartości rzeczywistej różni się o 0,456%. Spośród trzech analizowanych metod najlepsze wyniki przy sporządzaniu prognoz na tydzień w przód uzyskać można z metody naiwnej. Odmienne wnioski zostały otrzymane dla przewidywania wartości zapotrzebowania na energię na 2 i 3 tygodnie naprzód, dla których najlepszym sposobem okazał się wybór najbardziej podobnego dnia w 6-miesięcznej historii obserwacji. Jakością dopasowania w okresie czasu letniego i zimowego jest zgodna z wynikami uzyskanymi dla całej próby. Zaznaczyć jednak należy, że metoda dnia podobnego pozwala dokładniej szacować oczekiwaną wielkość zjawiskaw okresie obowiązywania czasu letniego. Uwzględnienie informacji o orkanach w Polsce nie pozwala na istotne poprawienie jakości dopasowania, gdyż wartości MAPE są na niemalże identycznym poziomie.

Wydaje się, że najrozsądniejszym sposobem pominięcia wpływu dni specjalnych na prognozy dla metody dnia podobnego jest usunięcie obserwacji za te dni z bazy danych. Dla najprostszej wersji procedury błąd MAPE spada o 0,1 punktu procentowego dla prognoz jednookresowych i praktycznie nie zmienia się dla pozostałych horyzontów. Większe spadki obserwować można dla dwóch pozostałych wariantów, ponieważ były one równe  $\frac{1}{3}$  i prawie  $\frac{1}{2}$  wartości błędów dla oryginalnych danych. Na podstawie uzyskanych wyników należałoby wyciągnąć wniosek, że w przypadku niedysponowania modelem ekonometrycznym najlepszą procedurą sporządzania prognoz byłoby wykorzystanie metody naiwnej do wyznaczania prognoz na 1 okres i metod dnia podobnego z przeszukiwaniem 6 poprzedzających miesięcy przy prognozach dwu- i trzyokresowych.

Pozornie zaskakujący wniosek o wyższości wyboru dnia najbardziej podobnego do prognozowanego spośród 6 poprzednich miesięcy może być wytłumaczony konkluzjami z rozdziałów 2 i 3. Obserwowana tendencja wzrostowa w zapotrzebowaniu na energię elektryczną i widoczny wpływ stanu gospodarki tłumaczą powyższy rezultat. Selekcja historycznej obserwacji o zbliżonych charakterystykach na podstawie całej próby może teoretycznie doprowadzić do wybrania obserwacji sprzed kilkunastu lat i z okresu o odwrotnym stanem koniunktury. Oznacza to, że jednym z kryteriów doboru najbardziej podobnej obserwacji powinno być podobieństwo fazy cyklu koniunkturalnego.

Tablica 5.2: Błędy MAPE dla różnych wersji metody dnia podobnego w %

	Ogółem			Czas zimowy			Czas letni		
	(1 dzień)	(2 dzień)	(3 dzień)	(1 dzień)	(2 dzień)	(3 dzień)	(1 dzień)	(2 dzień)	(3 dzień)
<i>Dane oryginalne</i>									
Metoda naiwna	0,456	0,875	1,167	0,532	0,886	1,156	0,383	0,864	1,179
Dzień podobny 6M	0,610	0,679	0,612	0,751	0,760	0,750	0,473	0,469	0,476
Dzień podobny	1,310	1,309	1,376	1,450	1,446	1,508	1,174	1,174	1,246
<i>Dane wygładzone</i>									
Metoda naiwna	0,459	0,875	1,166	0,539	0,887	1,152	0,383	0,864	1,179
Dzień podobny 6M	0,610	0,611	0,611	0,750	0,757	0,747	0,473	0,469	0,476
Dzień podobny	0,105	0,111	1,121	1,153	1,209	1,283	0,942	1,009	0,963
<i>Dane bez dni specjalnych</i>									
Metoda naiwna	0,358	0,815	1,121	0,369	0,792	1,078	0,347	0,835	1,162
Dzień podobny 6M	0,434	0,445	0,453	0,444	0,466	0,477	0,425	0,425	0,428
Dzień podobny	0,720	0,750	0,747	0,727	0,746	0,769	0,714	0,755	0,725

Horyzont prognozy na 1,2,3 następny dzień oznaczony jest jako (1 dzień), (2 dzień), (3 dzień).

Źródło: Opracowanie własne.



## 5.3 Model regresji

Pierwszym wyborem w modelowaniu, oczywiście, jeśli tylko może być zastosowany, jest standardowy model regresji. Ramanathan i inni<sup>15</sup> stwierdzili, że naturalną jest chęć wyestymowania jednego, skomplikowanego modelu dla całego szeregu czasowego. Model szacowany dla wszystkich obserwacji ma postać zbliżoną do modelu (4.13)-(4.14), z nieznaczną korektą dla funkcji określającej nieliniowy wpływ temperatury na zapotrzebowanie na energię, i przyjmuje postać

$$\begin{aligned} y_t = & \beta_0 + \beta_1 t + \beta_2 \text{godzina}_t + \beta_3 \text{dzien}_t + \beta_4 \text{miesiac}_t + \beta_5 \text{godzina}_t * \text{dzien}_t \\ & + \beta_6 t * \text{miesiac}_t + \beta_7 \text{lowcloct}_t + \beta_8 \text{sun}_t + \beta_9 \text{visibility}_t \\ & + \beta_{10} \text{dni\_specjalne}_t + \beta_{11} \text{temp\_med}_t + \beta_{12} \mathbf{1}_{(\text{temp\_med}_t > 16)} * \text{temp}_t + \varepsilon_t. \end{aligned} \quad (5.1)$$

Powyższa forma funkcyjna odpowiada modelowaniu funkcją ciągłą i przedziałami liniową.

Test Ramsey'a, nazywany też testem Reset, służy do badania poprawności formy funkcyjnej w modelu regresji. W teście Ramsey'a przeprowadza się regresję reszt  $e_t$  na zbiorze potęg wartości dopasowanych z modelu postaci

$$y_t = x_t \beta + \alpha_1 \hat{y}_t^2 + \dots + \alpha_p \hat{y}_t^{p+1} + \eta_t. \quad (5.2)$$

Weryfikuje się następnie hipotezę łączną  $H_0 : \alpha_1 = \dots = \alpha_p = 0$  za pomocą statystyki  $NR^2 \rightarrow \chi_p^2$ . Statystyka testowa uzyskana dla pełnego modelu oszacowanego na próbie od 1 stycznia 2002 r. do 10 października 2013 r. wyniosła 2608,75, przy 5% wartości krytycznej 2,605. Należy zatem odrzucić hipotezę zerową o poprawności formy funkcyjnej modelu. Jeden model dla wszystkich kwadransów, oprócz negatywnej weryfikacji testem Ramsey'a, nie spełnia wymogu poprawności formy funkcyjnej sprawdzanego także przy zastosowaniu testu Chowa. Test ten pozwala weryfikować hipotezę o stabilności parametrów w podpróbach. Konstrukcja testu pozwala analizować równość oszacowań w wielu podpróbach. Test Chowa do badania stabilności parametrów w wielu podpróbach nie jest zaprogramowany w programie Stata. Odpowiednie wyliczenia zostały wykonane samodzielnie. Do przeprowadzenia weryfikacji hipotezy o równości oszacowań dla wszystkich kwadransów niezbędne jest otrzymanie sumy kwadratów reszt dla modelu pełnego ( $S_R$ ) oraz dla modeli oszacowanych dla poszczególnych podprób, których suma oznaczana jest jako  $S$ . Statystyka testowa ma, w tym przypadku, postać

$$F = \frac{(S_R - S)/[K(m - 1)]}{S/(N - mK)} \sim F_{K(m-1), N-mK}, \quad (5.3)$$

gdzie  $K$  jest liczbą zmiennych w modelu ( $K = 130$ ),  $N = 412845$  - liczbą obserwacji, natomiast  $m$  - liczbą analizowanych podpróbek. Uzyskana wartość statystyki testowej  $F$

<sup>15</sup>Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-run Forecast...*, str. 165

wyniosła 53,68 i należała do obszaru krytycznego<sup>16</sup>, przez co należało odrzucić hipotezę o równości oszacowań we wszystkich podpróbach.

Odrzucenie hipotezy o równości nie jest jednoznaczne z tym, że parametry we wszystkich podpróbach są od siebie różne, a jedynie to, że dla przynajmniej jednej parametry są różne od pozostałych. Zdecydowałem, że kolejnym rozważanym modelem był model Engle’a, Grangera, Ramanathana i Vahida-Araghiego<sup>17</sup> (EGRV) w wersji uproszczonej<sup>18</sup>, który jest najbardziej ogólnym rozwiązaniem kwestii nierówności parametrów w podpróbach. Podejście to zakłada zbudowanie oddzielnego modelu regresji liniowej dla każdego z okresów w ciągu dnia. Ramanathan i inni zbudowali 24 osobne modele, każdy dla kolejnej godziny, ponieważ analizowali dane godzinowe. Dla rozważanych w niniejszej pracy danych kwadransowych należy zgodnie z tym oszacować 96 modeli, po jednym dla każdych 15 minut. Dla obserwacji od 1 stycznia 2002 do 10 października 2013 r. oszacowanych zostało 96 modeli o postaci zbliżonej do równania (5.1)-(5.2), pozbawionych jednak zmiennych dla efektu godzinowego. Ogólny model dla każdego kwadransa ma postać:

$$y_t = \beta_0 + \beta_1 t + \beta_2 \text{dzien}_t + \beta_3 \text{miesiac}_t + \beta_4 \text{dzien}_t * \text{miesiac}_t + \beta_5 t * \text{miesiac}_t + \beta_7 \text{lowclocta}_t + \beta_8 \text{sun}_t + \beta_9 \text{visibility}_t, \quad (5.4)$$

$$+ \beta_{10} \text{dni\_specjalne}_t + \beta_{11} \text{temp\_med} + \beta_{12} \mathbb{1}_{(\text{temp\_med}_t > 16)} * \text{temp}_t + \varepsilon_t. \quad (5.5)$$

Tablica 5.3 prezentuje oszacowania wybranych parametrów modeli szacowanych zgodnie z podejściem EGRV. Na jej podstawie można próbować wyciągnąć wniosek o odmienności oszacowań poszczególnych parametrów dla różnych kwadransów. Ramanathan i inni uzasadniają swoje podejście różnym wpływem poszczególnych zmiennych w różnych porach dnia<sup>19</sup>. Dzięki programowalnym termostatom wpływ temperatury w godzinach nocnych może być nieistotny, natomiast dla kwadransów porannych wpływ temperatury różny od zerowego. Niemniej istotne okazać się mogą wyniki dla poziomu zachmurzenia. Zauważalny jest wzrost wartości parametru dla godziny 12:00 i 17:00 w stosunku do godzin nocnych i godziny 7:00. Wyniki testu F wskazały na nierówność wpływu poziomu zachmurzenia na zapotrzebowanie na energię elektryczną dla poszczególnych godzin<sup>20</sup>. Analogicznie badanie wykonałem dla zmiennych  $\text{temp\_med}$  oraz  $\mathbb{1}_{(\text{temp\_med} > 16)}$  a z jego wyników wyciągnąłem wniosek o tym, że wpływ temperatury na popyt na energię nie jest równy dla wszystkich godzin<sup>21</sup>. Być może jest tak, że w Stanach Zjednoczonych obserwuje się różny wpływ temperatury na popyt na energię w różnych porach dnia, natomiast w Polsce wykorzystanie energii elektrycznej jest bardziej zróżnicowane w zależności od

<sup>16</sup>5% wartość krytyczna  $F_{0,05}(10830, 401901) \approx F_{0,05}(\infty, \infty) = 1,00$ .

<sup>17</sup>Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-Run Forecasts of Electricity Loads and Peaks*, International Journal of Forecasting, 13, s. 161-174.

<sup>18</sup>Ramanathan R., ..., str. 171.

<sup>19</sup>Ramanathan R., ..., str. 165

<sup>20</sup>Statystyka  $F(23, 412660)$  wyniosła 572,98, a odpowiadające jej p-value 0.

<sup>21</sup>Dla statystyki  $F(46, 412661)=746,56$  wartość p-value wyniosła 0.

zachmurzenia.

Tablica 5.3: Wybrane oszacowania modeli EGRV dla wybranych kwadransów

	01 : 00	07 : 00	12 : 00	17 : 00	22 : 00
t	0,000	0,000***	0,000***	0,000***	0,000***
temp_med	-0,007***	-0,006***	-0,006***	-0,006***	-0,006***
lowcloudocta	0,001*	0,002*	0,006***	0,007***	0,002***
sun	0,001**	0,001**	0,001***	0,001*	0,001*
visibility	-0,001**	-0,000	0,000	0,000	0,000**
Poniedziałek	0,012	0,251***	0,253***	0,207***	0,121***
Wtorek	0,107***	0,282***	0,261***	0,215***	0,131***
Środa	0,108***	0,284***	0,264***	0,213***	0,129***
Czwartek	0,116***	0,282***	0,260***	0,209***	0,125***
Piątek	0,113***	0,274***	0,253***	0,204***	0,116***
Sobota	0,088***	0,109***	0,138***	0,098***	0,034***
:	:	:	:	:	:
styczeń 1	-0,110***	-0,382***	-0,424***	-0,293***	-0,205***
styczeń 6	-0,035	-0,303***	-0,272***	-0,214***	-0,142***
maj 1	-0,060***	-0,261***	-0,232***	-0,259***	-0,155***
maj 3	-0,100***	-0,314***	-0,275***	-0,292***	-0,163***
listopad 11	-0,063***	-0,281***	-0,225***	-0,202***	-0,148***
długi weekend	-0,107***	-0,143***	-0,100***	-0,082***	-0,071***
$1_{(temp_t > 16)}$ temp_med	0,014***	0,011***	0,013***	0,014***	0,012***
Stała	9,606***	9,550***	9,597***	9,720***	9,703***
$R^2$	0,830	0,867	0,847	0,898	0,845
$R_a^2$	0,827	0,864	0,843	0,896	0,841
N	4747,000	4747,000	4748,000	4748,000	4748,000

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ 

Źródło: Opracowanie własne.

Oprócz modelu EGRV w wersji uproszczonej Ramanathan i inni<sup>22</sup> zaprezentowali wersję właściwą (rozszerzoną), uwzględniającą w modelowaniu błędy z poprzednich okresów. Przedstawiona przez Ramanathana i innych forma funkcyjna bardziej skomplikowanego modelu dla godziny 01:00 dla danych dla lat 1983-1990 ma postać<sup>23</sup>:

$$\begin{aligned}
 y_t = & 1678,9 + 40,6rok + 32,4grudzien + 0,5luty + 71,7marzec \\
 & - 193,9poniedzialek + 223,9dzien\_po\_swiecie \\
 & - 2,86temperatura + 0,39temperatura^2 \\
 & - 7,68\overline{temperatura}_7 - 3,62temperatura_{t-24} \\
 & + 0,08y_{08:00} - 0,10dzien\_po\_swiecie *_{08:00} \\
 & + 0,52\varepsilon_{t-1} + 0,15\varepsilon_{t-2} + 0,07\varepsilon_{t-3} + 0,14\varepsilon_{t-5}.
 \end{aligned} \tag{5.6}$$

Obecność opóźnionych reszt pozwala na precyzyjniejsze prognozowanie popytu na energię elektryczną. Doświadczony i krytyczny prognosta z pewnością uznałby korygowanie prognoz na podstawie przeszłych błędów za rozsądne<sup>24</sup>. Opóźnione reszty sprawiają, że model postaci 5.6 nie może być oszacowany standardową metodą najmniejszych kwadratów. Ramanathan i inni podali model regresji liniowej szacowany metodą najmniejszych kwadratów z przekształceniem Cochrane’a-Orcutta. Do porównań sposobów prognozowania polskich danych w pierwszej kolejności wykorzystany został model regresji z poprawką Cochrane’a-Orcutta na autoregresyjną strukturę błędu pierwszego rzędu<sup>25</sup>.

Regresja Cochrane’a-Orcutta<sup>26</sup> jest procedurą umożliwiającą uzyskanie zgodnych i nieobciążonych estymatorów szacowanych parametrów oraz macierzy wariancji-kowariancji w modelu z autoskorelowanymi resztami. W pierwszym jej etapie szacuje się wyjściowy model regresji

$$y_t = \beta_1 x_{1t} + \dots + \beta_K x_{Kt} + \varepsilon_t, \tag{5.7}$$

dla którego reszt wyznacza się współczynnik autokorelacji za pomocą regresji pomocniczej postaci

$$e_t = \rho e_{t-1} + u_t. \tag{5.8}$$

W drugim kroku stosuje się przekształcenie Cochrane’a-Orcutta

$$y_t - \rho y_{t-1} = \beta_1(x_{1t} - \rho x_{1t-1}) + \dots + \beta_K(x_{Kt} - \rho x_{Kt-1}) + \xi, \tag{5.9}$$

dla obserwacji  $t = 2, 3, \dots, T$ . Z oszacowanego równania (5.9) uzyskuje się wartości

<sup>22</sup>Ramanathan R., ..., str. 165

<sup>23</sup>Nazwy zmiennych zostały przetłumaczone na polski dla większej czytelności.

<sup>24</sup>Ramanathan R., 1993, ..., str. 164

<sup>25</sup>Taka wersja procedury była dostępna w programie Stata 11.2

<sup>26</sup>Cochrane D., Orcutt G.H., 1949, *Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms*, Journal of the American Statistical Association, 44(245), s. 32-61.

dopasowane, które w kroku trzecim służą do reestymacji parametru  $\rho$  w regresji

$$y_t - \hat{y}_t = \rho(y_{t-1} - \hat{y}_{t-1}) + \eta_t. \quad (5.10)$$

Uzyskane nowe oszacowanie wykorzystuje się do przekształcenia Cochrane'a-Orcutta i te dwa kroki powtarzane są aż do osiągnięcia zbieżności. Wartości parametru muszą zbiec do poszukiwanej wartości, a Sargan<sup>27</sup> zawiera tego dowód.

Postuluje się, między innymi Weron<sup>28</sup>, żeby prognozy z proponowanego modelu przechodziły pozytywnie test metody naiwnej. Wyniki uzyskane dla metod opartych o modele regresji liniowej zgodnie z tablicą 5.4 spełniają ten warunek. Model regresji dla wszystkich obserwacji charakteryzuje się niższym o jedną setną punktu procentowego błędem MAPE niż metoda naiwna dla prognozowania okres do przodu, natomiast dla dłuższego horyzontu prognozy różnica między tymi dwiema metodami jest większa. Zdecydowanie najdokładniejsza na całej próbie out-of-sample jest uproszczona wersja modelu EGRV. Ponownie, tak jak w podrozdziale 5.2, dla czasu zimowego uzyskiwane są mniej dokładne przewidywane wartości. Zaskakujący jest wniosek płynący z oszacowań modelu EGRV rozszerzonego o opóźnioną wartość błędu. Zastosowanie estymatora Cochrane'a-Orcutta skutkuje uzyskaniem mniej dopasowanych prognoz z modelu. Modele dla danych bez dni specjalnych wskazują na to, że spora część błędu prognozy związana jest z obecnością w bazie dni ustawowo wolnych od pracy. Zastosowana dla nich strategia polegała na wyestymowaniu modelu dla wszystkich obserwacji bez zmiennych zerojedynkowych dla świąt kościelnych i państwowych i zastąpieniem tych obserwacji wartościami dopasowanymi uzyskanymi z modelu.

Regresja Cochrane'a-Orcutta pozwala przezwyciężyć problem autokorelacji, ale niestety tylko pierwszego rzędu. Uwzględnienie autokorelacji reszt tylko pierwszego rzędu mogłoby być zakwestionowane w modelu dla danych 15-minutowych. Niezbędne dlatego jest zastosowanie modelu pozwalającego uwzględniać bardziej zaawansowaną strukturę autoregresyjną błędów.

Model (5.6) może być postrzegany jako szczególny przypadek modelu autoregresyjnego ze średnią ruchomą. W ramach tej klasy modeli możliwe jest także modelowanie z uwzględnieniem informacji spoza samego analizowanego szeregu. Podejście takie mogłoby być efektywniejsze i prostsze niż wykorzystywanie bardziej skomplikowanych narzędzi pozwalających uwzględniać autoregresyjną strukturę błędu.

---

<sup>27</sup>Sargan J. D., 1964, *Wages and Prices in the United Kingdom: A Study in Econometric Modelling*, w *Econometric Analysis for National Economic Planning*, ed. P.E. Hart, G. Mills, J.K. Whitaker, 25-64, London: Butterworths.

<sup>28</sup>Weron R., *Modelling and ...*, str. 64

Tablica 5.4: Błędy MAPE dla modeli regresji w %

	Ogółem			Czas zimowy			Czas letni		
	(1 dzień)	(2 dzień)	(3 dzień)	(1 dzień)	(2 dzień)	(3 dzień)	(1 dzień)	(2 dzień)	(3 dzień)
<i>Dane oryginalne</i>									
Model regresji	0,442	0,444	0,449	1,374	0,928	0,486	0,414	6,624	9,337
Uproszczony EGRV	0,342	0,347	0,352	0,407	0,412	0,419	0,281	0,284	0,286
Model EGRV(1)	0,392	0,398	0,403	0,482	0,489	0,496	0,305	0,309	0,312
<i>Dane wygładzone</i>									
Model regresji	0,527	0,530	0,534	0,568	0,573	0,580	0,486	0,488	0,489
Uproszczony EGRV	0,343	0,347	0,353	0,406	0,413	0,422	0,281	0,283	0,285
Model EGRV(1)	0,391	0,397	0,402	0,481	0,488	0,495	0,304	0,307	0,310
<i>Dane bez dni specjalnych</i>									
Model regresji	0,421	0,423	0,428	0,456	0,460	0,468	0,387	0,388	0,389
Uproszczony EGRV	0,264	0,269	0,274	0,301	0,307	0,315	0,229	0,231	0,234
Model EGRV(1)	0,297	0,301	0,306	0,320	0,325	0,333	0,275	0,278	0,280

Horyzont prognozy na 1,2,3 następny dzień oznaczony jest jako (1 dzień), (2 dzień), (3 dzień).

*Źródło:* Opracowanie własne.

## 5.4 Model autoregresyjny

Wymienione powody skłoniły mnie do oszacowania modeli będących rozszerzeniami modeli z poprzedniego podrozdziału. Uogólniona ich postać jest modelem ARMA ze zmiennymi egzogenicznymi (ARMAX) i jako jej podklasę wymienić można model ARDL. Generalna postać modeli tej kategorii to

$$y_t = x_t\beta + \rho_1(y_{t-1} - x_{t-1}\beta) + \rho_2(y_{t-2} - x_{t-2}\beta) + \dots + \rho_p(y_{t-p} - x_{t-p}\beta) + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} + \varepsilon_t, \quad (5.11)$$

gdzie  $\beta, \rho_1, \rho_2, \dots, \rho_p, \theta_1, \theta_2, \dots, \theta_q$  są nieznanymi wartościami parametrów,  $\varepsilon_t$  jest błędem losowym z okresu  $t$ , natomiast  $\varepsilon_{t-i}$  jest błędem losowym sprzed  $i$  okresów. Dwoma pierwszymi zaprezentowanymi modelami są modele EGRV rozszerzone o opóźnione błędy losowe, estymowane jako procesy ARMAX, dla całej próby i próbki 2008-2013. Trzecim wariantem był model ARMA bez średniej ruchomej, czyli ARDL, którego postać otrzymuje się z (5.11) zakładając  $\theta_1 = \theta_2 = \dots = \theta_q = 0$ . Ze względu na niestacjonarność szeregu zapotrzebowania na moce przesyłowe odjęto od niego deterministyczny trend.

Modeli ARMA są często wykorzystywane w literaturze do modelowania i prognozowania popytu na energię elektryczną. Wymienić dla przykładu można prace Nowickiej-Zagrajek i Weron<sup>29</sup>, Misiorka i Weron<sup>30</sup>, natomiast dla polskich danych prace Szmit<sup>31</sup> i Lichoty<sup>32</sup>.

Tablica 5.5 zawiera wyniki eksperymentu opisanego w podrozdziale 5.1 dla tych modeli. Pierwszym wnioskiem jest pozytywny wpływ rozszerzenia modeli EGRV o opóźnione wartości błędów na dokładność prognoz. Modele te są jednak o rząd wielkości mniej precyzyjne niż jeden model ARDL oszacowany dla całej próby, bez podziałów na kwadranse. Podjęte inne uzyskania dokładniejszych prognoz niż z modelu ARDL nie przyniosły zamierzonego rezultatu. Dla przykładu błąd MAPE dla modelu autoregresyjnego średniej ruchomej dla poszczególnych kwadransów został zredukowany o około 0.04 punktu procentowego, ale i tak był trzykrotnie wyższy niż uzyskany z modelu ARDL. Dla modelu autoregresyjnego ze zmiennymi egzogenicznymi, jak i dla pozostałych, obserwuje się nieintuicyjny wzrost wartości wskaźnika MAPE dla skorygowanych danych zarówno o informacje o orkanach, jak i o dniach specjalnych. Dodatkowy komentarz do modelu ARDL znajduje się w podrozdziale 5.6.

<sup>29</sup>Nowicka-Zagrajek J., Weron R., 2002, *Modeling Electricity Loads in California: ARMA Models with Hyperbolic Noise*, Signal Processing 82, s. 1903-1915.

<sup>30</sup>Misiorek A., Weron R., 2004, *Modelowanie sezonowości a prognozowanie zapotrzebowania na energię elektryczną*, Energetyka 12/2004, s. 794-799.

<sup>31</sup>Szmit A., 2002, *Prognozowanie zapotrzebowania na energię elektryczną. Studium empiryczne dla regionu łódzkiego*, Rozprawa doktorska, Politechnika Łódzka, Łódź.

<sup>32</sup>Lichota A., 2006, *Prognozowanie krótkoterminowe na lokalnym rynku energii elektrycznej*, Rozprawa doktorska, Akademia Górniczo-Hutnicza, Kraków, s. 176.



Tablica 5.5: Błędy MAPE dla modeli klasy ARMA

	Ogółem			Czas zimowy			Czas letni		
	(1 dzień)	(2 dzień)	(3 v)	(1 dzień)	(2 dzień)	(3 dzień)	(1 dzień)	(2 dzień)	(3 dzień)
<i>Dane oryginalne</i>									
EGRV+MA(1,7) rok>2008	0,292	0,294	0,297	0,346	0,348	0,353	0,240	0,241	0,242
EGRV+MA(1,7) ∇obs,	0,285	0,286	0,289	0,310	0,312	0,317	0,260	0,261	0,261
Model ARDL	0,080	0,080	0,080	0,083	0,083	0,083	0,077	0,077	0,077
<i>Dane wygładzone</i>									
EGRV+MA(1,7) rok>2008	0,276	0,275	0,278	0,317	0,311	0,312	0,236	0,241	0,244
EGRV+MA(1,7) ∇obs,	0,297	0,296	0,298	0,329	0,323	0,324	0,264	0,269	0,273
Model ARDL	0,084	0,084	0,084	0,085	0,086	0,086	0,083	0,083	0,083
<i>Dane bez dni specjalnych</i>									
EGRV+MA(1,7) rok>2008	0,298	0,297	0,300	0,358	0,352	0,352	0,239	0,244	0,248
EGRV+MA(1,7) ∇obs,	0,327	0,327	0,329	0,386	0,380	0,380	0,270	0,275	0,278
Model ARDL	0,084	0,084	0,084	0,085	0,086	0,086	0,083	0,083	0,083

Źródło: Opracowanie własne.

## 5.5 Model splajnów

Niniejsza sekcja przedstawia splajny stopnia trzeciego oraz sposoby, jak mogą one zostać użyte w modelach regresji i szeregów czasowych. Zaprezentowałem teorię dość szczegółowo, ponieważ w dostępnych opracowaniach nie wszystkie wyprowadzenia są przejrzyste. Podrozdział 5.5.1, zawiera więcej algebraicznych przekształceń niż ekonometrycznych, natomiast w podrozdziałach 5.5.2-5.5.4 wykorzystana została przedstawiona teoria w zastosowaniach ekonometrycznych.

### 5.5.1 Model splajnów kubicznych

Odpowiedzią na pytanie o przyczynę popularności splajnów trzeciego stopnia wydaje się być kompromis między gładkością wyników i względnym nieskomplikowaniem metody. Narzędzie to jest powszechnie wykorzystywane w wielu dziedzinach nauk i w praktyce. Niniejszy podrozdział zawiera opis splajnów trzeciego stopnia ze światowej i z polskiej literatury.

Niech  $\Delta = \{\bar{x}_0 < \bar{x}_1 < \dots < \bar{x}_k\}$  będzie siatką punktów na przedziale  $[\bar{x}_0, \bar{x}_k]$ , dla  $k+1 \geq 3$ , natomiast poszczególne punkty  $\bar{x}_j$ , dla  $j = 0, 1, \dots, k$  noszą nazwę węzłów. Niech zbiór  $y = \{y_0, y_1, \dots, y_k\}$  będzie odpowiadającym im zbiorem wartości, wówczas splajnem sześciennym (kubicznym) na  $\Delta$  interpolującym  $y$ , oznaczanym jako  $S_\Delta(x)$  jest funkcja spełniająca warunki:

- a)  $S_\Delta(x)$  i jej dwie pierwsze pochodne są ciągłe w  $[\bar{x}_0, \bar{x}_k]$ ,
- b)  $S_\Delta(x)$  jest wielomianem stopnia co najwyżej trzeciego w każdym z podprzedziałów  $[\bar{x}_{j-1}, \bar{x}_j]$  dla  $j = 1, 2, \dots, k$ ,
- c)  $S_\Delta(\bar{x}_j) = y_j$  dla wszystkich  $j$ .

Wykorzystując równanie dla funkcji liniowej dla przedziału  $[a, b]$

$$f(x) = f(a) \frac{b-x}{b-a} + f(b) \frac{x-a}{b-a} \quad (5.12)$$

i podstawiając  $f(x) = S''_\Delta(x)$ ,  $M_{j-1} = S''_\Delta(\bar{x}_{j-1})$ ,  $M_j = S''_\Delta(\bar{x}_j)$  otrzymuje się

$$S''_\Delta(x) = \left[ \frac{\bar{x}_j - x}{h_j} \right] M_{j-1} + \left[ \frac{x - \bar{x}_{j-1}}{h_j} \right] M_j, \quad (5.13)$$

gdzie  $h_j$  jest długością  $j$ -tego podprzedziału, to jest  $h_j = \bar{x}_j - \bar{x}_{j-1}$ . Całkując dwukrotnie

równanie 5.13 uzyskuje się

$$S_{\Delta}(x) = \frac{\bar{x}_j - x}{6h_j} [(\bar{x}_j - x)^2 - h_j^2] M_{j-1} + \frac{x - \bar{x}_{j-1}}{6h_j} [(x - \bar{x}_{j-1})^2 - h_j^2] M_j \\ + \left[ \frac{\bar{x}_j - x}{h_j} \right] y_{j-1} + \left[ \frac{x - \bar{x}_{j-1}}{h_j} \right] y_j. \quad (5.14)$$

Pokazać można, że warunki (a)-(c) determinują układ równań

$$(1 - \lambda_j) M_{j-1} + 2M_j + \lambda_j M_{j+1} = \frac{6y_{j-1}}{h_j(h_j + h_{j+1})} + \frac{6y_j}{h_j h_{j+1}} + \frac{6y_{j+1}}{h_{j+1}(h_j + h_{j+1})}, \quad (5.15)$$

gdzie  $\lambda_j = \frac{h_{j+1}}{h_j + h_{j+1}}$  dla  $j = 1, 2, \dots, k-1$ . Wyprowadzenie powyższe zostało wyjaśnione m. in. w Poirier<sup>33</sup> oraz na przykład w polskojęzycznej literaturze w Fortuna, Macukow i Wąsowski<sup>34</sup>. Zauważyć należy, że na podstawie 5.15 otrzymuje się  $k-1$  równań, więc do istnienia jednoznacznego rozwiązania potrzeba dodatkowych dwóch. W Fortuna, Macukow i Wąsowski<sup>35</sup> stwierdzone zostało, że najczęściej do układu równań 5.15 dokłada się ograniczenie na  $S'(x_l)$  lub  $S''(x_l)$  dla  $l = 0, k$ . Rozważmy ograniczenia końcowe (brzegowe) na drugą pochodną  $S''(\bar{x}_0) = 0$  i  $S''(\bar{x}_k) = 0$ , tak jak w Poirier<sup>36</sup>. Przy zdefiniowanych macierzach

$$\Lambda = \begin{bmatrix} 2 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 - \lambda_1 & 2 & \lambda_1 & \dots & 0 & 0 & 0 \\ 0 & 1 - \lambda_2 & 2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & \lambda_{k-2} & 0 \\ 0 & 0 & 0 & \dots & 1 - \lambda_{k-1} & 2 & \lambda_{k-1} \\ 0 & 0 & 0 & \dots & 0 & 0 & 2 \end{bmatrix} \quad (5.16)$$

$$\Theta = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{6}{h_1(h_1+h_2)} & \frac{-6}{h_1 h_2} & \frac{6}{h_2(h_1+h_2)} & \dots & 0 & 0 & 0 \\ 0 & \frac{6}{h_2(h_2+h_3)} & \frac{-6}{h_2 h_3} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{-6}{h_{k-2} h_{k-1}} & \frac{6}{h_{k-1}(h_{k-2}+h_{k-1})} & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix} \quad (5.17)$$

oraz wektorach

$$M = [M_0, M_1, \dots, M_k]', \quad (5.18)$$

$$y = [y_0, y_1, \dots, y_k]', \quad (5.19)$$

<sup>33</sup>Poirier D.J. (1976), *The Econometrics of Structural Change: With Special Emphasis on Spline Functions*, Amsterdam: North-Holland, str. 21-55.

<sup>34</sup>Fortuna Z., Macukow B., Wąsowski J. (2001), *Metody numeryczne*, Wydawnictwa Naukowo Techniczne, Warszawa, s. 383.

<sup>35</sup>Fortuna Z., Macukow B., Wąsowski J. (2001), *Metody numeryczne*, ..., str. 66.

<sup>36</sup>Poirier D.J. (1976), *The Econometrics of Structural Change...*

równania 5.15 można zapisać w postaci macierzowej

$$\Lambda M = \Theta y. \quad (5.20)$$

Dla dowolnego wektora argumentów  $\xi = [\xi_1, \xi_2, \dots, \xi_n]'$  i odpowiadającego mu wektora  $S_\Delta(\xi) = [S_\Delta(\xi_1), S_\Delta(\xi_2), \dots, S_\Delta(\xi_n)]'$  interpolujących wartości ze splajnu można przedstawić jako liniową kombinację wektora wartości  $y$ . W celu przedstawienia wartości interpolujących dla punktów  $\xi$ , w zależności od wektora  $y$ , należy zdefiniować najpierw dwie macierze  $P = [p_{im}]$  oraz  $Q = [q_{im}]$ . Dla  $\bar{x}_{j-1} \leq \xi \leq \bar{x}_j$  i dla  $i = 1, 2, 3, \dots, n$  mają one postać

$$p_{im} = \begin{cases} (\bar{x}_j - \xi_i) [(\bar{x}_j - \xi_i)^2 - h_j^2] / 6h_j, & \text{dla } m = j - 1 \\ (\xi_i - \bar{x}_{j-1}) [(\xi_i - \bar{x}_{j-1})^2 - h_j^2] / 6h_j, & \text{dla } m = j \\ 0, & \text{w p.p.} \end{cases} \quad (5.21)$$

$$q_{im} = \begin{cases} (\bar{x}_j - \xi_i) / h_j, & \text{dla } m = j - 1 \\ (\xi_i - \bar{x}_{j-1}) / h_j, & \text{dla } m = j \\ 0, & \text{w p.p.} \end{cases} \quad (5.22)$$

Korzystając z (5.21), (5.22) oraz (5.14) otrzymuje się

$$\begin{aligned} S_\Delta &= PM + QY \\ &= (P\Lambda^{-1}\Theta + Q)Y \\ &= WY. \end{aligned} \quad (5.23)$$

Przedstawiona metoda wydaje się być idealna do interpolacji określonych punktów węzłowych. Taylor i Buizza<sup>37</sup> opisali sposób wyznaczania prognoz przez National Grid - operatora sieci elektroenergetycznych w Wielkiej Brytanii, który polega na wyznaczeniu wartości dla kilku punktów w ciągu dnia i interpolacji tych wartości. Zagadnienie to sprowadza się do interpolacji splajnami kubicznymi wybranych węzłów. Procedura taka zakłada dopasowanie do ustalonych punktów a dopasowanie funkcji w punktach niewęzłowych nie jest uwzględniane. Zdefiniować można dlatego regresję splajnami sześciennymi.

---

<sup>37</sup>Taylor J.W., Buizza R., 2003, *Using Weather Ensemble Predictions in Electricity Demand Forecasting*, International Journal of Forecasting, 19, s. 57-70.

### 5.5.2 Regresja splajnami sześciennymi

Model regresji najmniejszych kwadratów splajnami kubicznymi określony jest dla siatki punktów  $\Delta$ , dla których wektor wartości  $y$ , odpowiadający węzłom z  $\Delta$ , jest nieznanym. Niech  $\xi = [\xi_1, \xi_2, \dots, \xi_n]'$  i  $\eta = [\eta_1, \eta_2, \dots, \eta_n]'$  będą wektorami  $n \geq k + 1$  wartości<sup>38</sup> odpowiednio zmiennej niezależnej  $\xi$  i zależnej  $\eta$ . Ponadto, niech  $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]'$  będzie wektorem niezależnych zaburzeń losowych o jednakowym rozkładzie normalnym, takim, że  $E(\varepsilon) = 0$  oraz  $E(\varepsilon'\varepsilon) = \sigma^2 I$ . Model regresji splajnami kubicznymi określony dla splajnu sześciennego  $S_\Delta(\xi) = Wy$  ma postać

$$\eta = S_\Delta(\xi) + \varepsilon = Wy + \varepsilon. \quad (5.24)$$

Najlepiej dopasowany splajn pod względem kryterium błędu średniokwadratowego to

$$\hat{S}_\Delta(\xi) = W\hat{y}, \quad (5.25)$$

gdzie  $\hat{y}$  można przedstawić, przy użyciu standardowego wzoru regresji MNK ( $\hat{\beta} = (X'X)^{-1}X'y$ ), jako

$$\hat{y} = (W'W)^{-1}W'\eta. \quad (5.26)$$

Dla danych 15-minutowych o zapotrzebowaniu na moce przesyłowe KSE należałoby tak wybrać węzły siatki, żeby za ich pomocą jak najlepiej można było modelować skomplikowaną strukturę sezonowości szeregu. Literatura przedmiotu, na przykład Poirier<sup>39</sup>, zawiera rozszerzenie techniki splajnów na przypadki zmiennych sezonowych.

### 5.5.3 Okresowe splajny sześciennie

Kubiczny splajn okresowy zdefiniowany na  $[\bar{x}_0, \bar{x}_k]$  jest zdefiniowany, jak na przykład w Poirier<sup>40</sup>, poprzez warunki (a)–(c) z podrozdziału 5.5.1 z dodatkowym warunkiem postaci

$$S_\Delta^{(m)}(\bar{x}_0) = S_\Delta^{(m)}(\bar{x}_k), \quad m = 0, 1, 2. \quad (5.27)$$

Warunek ten jest dosyć intuicyjny, ponieważ nałożone ograniczenie można interpretować w ten sposób, że okresowy splajn powinien zaczynać i kończyć się w ten sam sposób. Równanie 5.15 dla  $k$ -tego podprzedziału zapisać można w postaci

$$(1 - \lambda_k) M_{k-1} + 2M_k + \lambda_k M_1 = \frac{6y_{k-1}}{h_k(h_k + h_1)} - \frac{6y_k}{h_k h_1} + \frac{6y_1}{h_1(h_k + h_1)}, \quad (5.28)$$

<sup>38</sup>  $k + 1$  równe jest liczbie punktów siatki  $\Delta$ .

<sup>39</sup> Poirier D.J. (1976), *The Econometrics of Structural Change...*, str. 43-47

<sup>40</sup> Poirier D.J. (1976), *The Econometrics of Structural Change...*, str. 44

gdzie  $\lambda_k = \frac{h_1}{h_k + h_1}$ . Analogicznie do (5.16) i (5.17) zdefiniować można ich odpowiedniki dla splajnu okresowego wymiaru  $k \times k$

$$\tilde{\Lambda} = \begin{bmatrix} 2 & \lambda_1 & 0 & \dots & 0 & 0 & 1 - \lambda_1 \\ 1 - \lambda_2 & 2 & \lambda_1 & \dots & 0 & 0 & 0 \\ 0 & 1 - \lambda_3 & 2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & 1 - \lambda_{k-2} & 0 \\ 0 & 0 & 0 & \dots & 1 - \lambda_{k-1} & 2 & \lambda_{k-1} \\ \lambda_k & 0 & 0 & \dots & 0 & 1 - \lambda_k & 2 \end{bmatrix}, \quad (5.29)$$

$$\tilde{\Theta} = \begin{bmatrix} \frac{6}{h_1 h_1} & \frac{6}{h_2(h_1+h_2)} & 0 & \dots & 0 & 0 & \frac{6}{h_1(h_2+h_2)} \\ \frac{6}{h_2(h_2+h_3)} & \frac{-6}{h_2 h_3} & \frac{6}{h_3(h_2+h_3)} & \dots & 0 & 0 & 0 \\ 0 & \frac{6}{h_3(h_3+h_4)} & \frac{-6}{h_3 h_4} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{-6}{h_{k-2} h_{k-1}} & \frac{6}{h_{k-1}(h_{k-2}+h_{k-1})} & 0 \\ 0 & 0 & 0 & \dots & \frac{6}{h_{k-1}(h_{k-1}+h_k)} & \frac{-6}{h_{k-1} h_k} & \frac{6}{h_k(h_{k-1}+h_k)} \\ \frac{6}{h_1(h_k+h_1)} & 0 & 0 & \dots & 0 & \frac{6}{h_k(h_k+h_1)} & \frac{-6}{h_k h_1} \end{bmatrix}, \quad (5.30)$$

natomiast odpowiednikami  $M$  oraz  $y$  są

$$\begin{aligned} \tilde{M} &= [M_1, M_2, \dots, M_k]', \\ \tilde{y} &= [y_1, y_2, \dots, y_k]'. \end{aligned} \quad (5.31)$$

Wówczas warunek (5.20) dla splajnów okresowych ma postać

$$\tilde{\Lambda} \tilde{M} = \tilde{\Theta} \tilde{y}. \quad (5.32)$$

Postać modelu regresji najmniejszych kwadratów okresowymi splajnami kubicznymi dla wcześniej zdefiniowanego wektora współrzędnych  $\xi$  oraz odpowiadającego mu wektora wartości  $\eta$  ma postać

$$\begin{aligned} S_{\Delta} &= \tilde{P} \tilde{M} + \tilde{Q} \tilde{y} \\ &= \left( \tilde{P} \tilde{\Lambda}^{-1} \tilde{\Theta} + \tilde{Q} \right) \tilde{y} \\ &= \tilde{W} \tilde{y}, \end{aligned} \quad (5.33)$$

gdzie macierze  $\tilde{P} = [\tilde{p}_{im}]$  i  $\tilde{Q} = [\tilde{q}_{im}]$  są zdefiniowane analogicznie do (5.21)-(5.22), to jest dla  $j = 2, 3, \dots, k$

$$\tilde{p}_{im} = \begin{cases} (\bar{x}_j - \xi_i) [(\bar{x}_j - \xi_i)^2 - h_j^2] / 6h_j, & \text{dla } m = j - 1 \\ (\xi_i - \bar{x}_{j-1}) [(\xi_i - \bar{x}_{j-1})^2 - h_j^2] / 6h_j, & \text{dla } m = j \\ 0, & \text{w p.p.} \end{cases}, \quad (5.34)$$

$$\tilde{q}_{im} = \begin{cases} (\bar{x}_j - \xi_i) / h_j, & \text{dla } m = j - 1 \\ (\xi_i - \bar{x}_{j-1}) / h_j, & \text{dla } m = j \\ 0, & \text{w p.p.} \end{cases}, \quad (5.35)$$

oraz dla  $j = 1$  określone jako

$$\tilde{p}_{im} = \begin{cases} (\bar{x}_j - \xi_i) [(\bar{x}_j - \xi_i)^2 - h_j^2] / 6h_j, & \text{dla } m = k \\ (\xi_i - \bar{x}_{j-1}) [(\xi_i - \bar{x}_{j-1})^2 - h_j^2] / 6h_j, & \text{dla } m = 1 \\ 0, & \text{w p. p.} \end{cases}, \quad (5.36)$$

$$\tilde{q}_{im} = \begin{cases} (\bar{x}_j - \xi_i) / h_j, & \text{dla } m = k \\ (\xi_i - \bar{x}_{j-1}) / h_j, & \text{dla } m = 1 \\ 0, & \text{w p. p.} \end{cases}. \quad (5.37)$$

#### 5.5.4 Specyfikacja modelu i wyniki

Harvey i Koopman<sup>41</sup> zaprezentowali dwie wersje modelu zmiennych w czasie splajnów sześciennych, z okresami jednego dnia i jednego tygodnia. Preferowany jest wybór okresu tygodniowego jako długości funkcji sklejaney<sup>42</sup>. W niniejszym opracowaniu wykorzystana została uproszczona wersja metody regresji stałymi w czasie splajnami kubicznymi. Argumentami za uproszczeniem modelu były: czas estymacji, deterministyczną wersją splajnów oraz możliwość estymacji metodą najmniejszych kwadratów.

Model regresji najmniejszych kwadratów splajnami kubicznymi użyty do prognozowania krótkookresowego zapotrzebowania miał postać

$$y_t = \beta_0 + \beta_1 t + \beta_4 temp_t + \beta_5 \mathbf{1}_{temp > 16} temp_t + \psi_t + \gamma_t + \sum_{i=1}^k \delta_i \chi_{it} + \varepsilon_t, \quad (5.38)$$

gdzie  $t = 1, 2, \dots, T$ . W równaniu (5.38) zmienna  $\psi_t$  modeluje sezonowość roczną i jest zdefiniowana jako

$$\psi_t = \alpha_1 \cos(\omega t) + \alpha_2 \sin(\omega t) \quad (5.39)$$

dla  $\omega = \frac{2\pi}{35040}$ , komponent  $\gamma_t$  jest tygodniowym splajnem kubicznym, natomiast  $\chi_{it}$  jest wartością  $i$ -tej pogodowej zmiennej egzogenicznej w okresie  $t$ . Elementy splajnu  $\gamma$  zostały

<sup>41</sup>Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity Demand...*

<sup>42</sup>Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity Demand...*, str. 1230-1231.

wyznaczone zgodnie z (5.33), w którym  $\widetilde{W} = (\widetilde{P}\widetilde{\Lambda}^{-1}\widetilde{\Theta} + \widetilde{Q})$  z pominięciem jej ostatniej kolumny<sup>43</sup>. Wyznaczonych zostało 6 różnych splajnów tygodniowych, po jednym dla każdej pory roku oraz dla zmian czasu na wiosnę i jesienią, co wynika z niestandardowej liczby kwadransów dla tych tygodni. Różna struktura zużycia energii odpowiadająca porom roku była powodem, dla którego wykorzystane zostały różne postaci splajnu. Analiza graficzna i proste próby dopasowania funkcji sklepanych trzeciego stopnia w poszczególnych porach roku wskazują jednoznacznie na potrzebę odrębnego modelowania przynajmniej czterech przypadków. Niejednolite rozkłady wykorzystania energii elektrycznej implikują odmienny wybór punktów węzłowych, które zostały zaprezentowane w tablicy 5.6.

Tablica 5.6: Punkty węzłowe  $\Delta$  w regresji splajnami kubicznymi

Pora roku	Węzły						
	Dzień roboczy				Weekend		
Wiosna	02:00	05:45	07:15	08:45	02:00	05:45	09:15
	10:15	12:00	14:00	17:30	12:00	16:45	19:30
	18:15	21:00	22:15	23:30	21:00	22:15	23:30
Lato	00:15	02:00	03:15	05:30	02:00	05:45	11:30
	09:00	12:00	13:15	14:00	13:15	14:00	18:00
	16:45	19:30	21:30	23:15	21:15	22:25	23:30
Jesień	02:00	04:00	05:30	08:30	02:00	04:00	06:15
	09:30	10:15	12:00	13:30	12:30	13:30	16:15
	17:30	19:30	22:00	23:30	18:30	19:30	22:00
Zima	03:45	05:30	07:15	07:45	03:45	05:30	07:15
	10:00	12:00	13:15	15:30	07:45	13:00	15:15
	16:30	18:15	19:30	22:00	18:15	19:30	22:00

*Źródło:* Opracowanie własne.

Model regresji okresowymi splajnami kubicznymi został zaprogramowany w środowisku MatLab i ze względu na moc obliczeniową posiadanego komputera wybrałem 77 punktów węzłowych dla każdego tygodnia.

Wyniki eksperymentu prognostycznego dla modelu splajnów wskazują na słabe dopasowanie w okresie letnim, co jest odmiennym wnioskiem do wszystkich prezentowanych wyżej sposobów prognozowania. Wartości błędów MAPE dla całego zbioru danych poza próbą są bardzo zbliżone dla trzech horyzontów prognoz. Podobną cechą charakteryzują się, z przedstawionych, jedynie prognozy z modelu ARDL. Odmienne uwzględnienie w modelu

<sup>43</sup>Poirier D.J., 1976, *The Econometrics of Structural Change*, ..., str. 47 zauważył, że usunięcie jednej z kolumn macierzy  $\widetilde{W}$  jest niezbędne w modelu z innymi zmiennymi egzogenicznymi ze względu na problem współliniowości ze stałą.



dni orkanów oraz świąt kościelnych i państwowych przekłada się na zmniejszenie błędu prognoz. Zauważyć należy, że wartość średniego błędu dla następnego kwadransów kolejnego dnia była równa 0,851%, przy analogicznej mierze 0,456% dla prognoz z najprostszej wersji metody naiwnej<sup>44</sup>. Prognozy z modelu splajnów nie przechodzą testu prognoz naiwnych. Wysokie wartości średnich względnych błędów przewidywania wynikać mogą z kilku powodów. Po pierwsze, spowodowane to być może zbyt małą liczbą punktów węzłowych, po drugie, prawdopodobnie potrzebne jest zaproponowanie osobnych splajnów tygodniowych dla krótszych niż pory roków okresów (np.: dla miesięcy) a po trzecie, z bardziej złożonej postaci sezonowości rocznej  $\psi$ .

Tablica 5.7: Błędy MAPE dla modelu okresowych splajnów kubicznych w %

	Ogółem			Czas zimowy			Czas letni		
	(1d)	(2d)	(3d)	(1d)	(2d)	(3d)	(1d)	(2d)	(3d)
<i>Dane oryginalne</i>									
Splajny okresowe	0,851	0,853	0,853	0,717	0,721	0,721	1,86	2,3	2,7
<i>Dane wygładzone</i>									
Splajny okresowe	0,839	0,841	0,840	0,700	0,701	0,701	1,847	2,284	2,720
<i>Dane bez dni specjalnych</i>									
Splajny okresowe	0,768	0,770	0,769	0,590	0,591	0,591	1,814	2,251	2,687

1d, 2d, 3d oznaczają horyzonty prognozy odpowiednio na 1, 2 i 3 dni.

Źródło: Opracowanie własne.

<sup>44</sup>Również porównując błędy MAPE dla horyzontów 2 i 3 dni stwierdzić należy, że z modelu naiwnego otrzymać można dokładniejsze prognozy.

## 5.6 Prognozowanie krótkookresowe

Cztery sposoby prognozowania przedstawione we wcześniejszych podrozdziałach zestawilem w tablicy 5.8. Wyniki dla metody dnia podobnego złożone zostały z metody naiwnej, dla horyzontu jednodniowego, i przeszukiwania historii sześciomiesięcznej, dla pozostałych dwóch horyzontów, co uzasadnione jest wartościami błędów MAPE. Drugim wyborem był najlepszy spośród modeli EGRV, czyli ten w uproszczonej, pozbawionej opóźnionej wartości błędu, postaci. Klasę ARIMA reprezentuje natomiast model ARDL.

Porównanie modeli dla prognozowania krótkookresowego wypada niezwykle korzystnie dla najprostszego modelu ARDL. Uwzględnienie jednodniowego i jednotygodniowego opóźnienia pozwala otrzymać model precyzyjniejszy o rząd wielkości w wartościach średnich względnych błędów prognozy. Model regresji okresowymi splajnami kubicznymi okazał się najmniej dokładny, Hierarchia modeli jest zachowana w przypadku prognozowania na skorygowanych danych o dni orkanów i dni specjalne. Na podstawie uzyskanych wartości błędów MAPE odrzucić należy pomocniczą hipotezę o wyższości sposobu modelowania zaproponowanego przez Ramanathan i innych<sup>45</sup> dla polskich danych.

W najdokładniejszym modelu na podstawie wartości odpowiednich statystyk testowych odrzucić należy założenie o braku autokorelacji i poprawności formy funkcyjnej. Problem skorelowanych reszt można rozwiązać poprzez zastosowanie odpornej macierzy Newey’a-Westa, natomiast drugi problem nie może być rozwiązany inaczej niż przez dodanie zmiennych, interakcji lub nieliniowych zależności. Pokazać można, że modele klasy EGRV pozytywnie weryfikowane są testami diagnostycznymi, tak jak model przedstawiony w rozdziale 4.3. Z punktu widzenia praktyki lepszy może być model dokładniejszy, nawet pomimo niespełnienia założeń. Niejednoznaczna i niezwykle interesującą kwestią jest analiza teoretyczna niniejszego wyboru. W koncepcjach ekonomicznych postuluje się konieczność zbudowania jak najprostszego modelu opisującego zjawisko, w tym przypadku jeden model regresji dla wszystkich kwadransów spełnia ten wymóg. Wnioskowane wymaganie najdokładniejszego charakteryzowania zmiennej może nie być spełnione w przypadku obciążonych i niezgodnych estymatorów parametrów i macierzy wariancji-kowariancji.

Niewątpliwie, przyszłe analizy powinny uwzględniać powyższy dylemat. Udoskonalic należy podejście wykorzystujące oddzielne modelowanie poszczególnych kwadransów, tak, by móc w prognozowaniu wartości dla danego kwadransa móc odwoływać się do większej liczby opóźnień błędów, w tym także do tych dla innych kwadransów. Wyraźnie wskazana nieefektywność regresji okresowymi splajnami sześciennymi może być poprawiona dzięki zastosowaniu komputera z większą mocą obliczeniową lub, co bardziej ekonomiczne, estymacji modelu za pomocą filtru Kalmana i uzmiennieniu w czasie funkcji sklepanych. Nadmienię pragnę, że przedstawione wyniki uzyskałem przy założeniu niewykorzystywa-

<sup>45</sup>Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-Run Forecasts...*

nia informacji o cyklu koniunkturalnym zgodnie z argumentacją Ramanathana i innych<sup>46</sup>.

Tablica 5.8: Porównanie błędów MAPE dla modeli dla prognoz krótkookresowych w %

	Ogółem			Czas zimowy			Czas letni		
	(1d)	(2d)	(3d)	(1d)	(2d)	(3d)	(1d)	(2d)	(3d)
<i>Dane oryginalne</i>									
Metoda naiwna	0,456	0,679	0,612	0,532	0,760	0,750	0,383	0,469	0,476
Model EGRV	0,342	0,347	0,352	0,407	0,412	0,419	0,281	0,284	0,286
Model ARDL	0,080	0,080	0,080	0,083	0,083	0,083	0,077	0,077	0,077
Splajny okresowe	0,851	0,717	0,853	0,717	0,721	0,721	1,86	2,3	2,7
<i>Dane wygładzone</i>									
Metoda naiwna	0,459	0,611	0,611	0,539	0,757	0,747	0,313	0,469	0,476
Model EGRV	0,343	0,347	0,353	0,406	0,413	0,422	0,281	0,283	0,285
Model ARDL	0,084	0,084	0,084	0,085	0,085	0,086	0,083	0,083	0,083
Splajny okresowe	0,839	0,841	0,840	0,700	0,701	0,701	1,847	2,284	2,720
<i>Dane bez dni specjalnych</i>									
Metoda naiwna	0,358	0,445	0,453	0,369	0,466	0,477	0,347	0,425	0,428
Model EGRV	0,264	0,269	0,274	0,301	0,307	0,315	0,229	0,231	0,234
Model ARDL	0,084	0,084	0,084	0,085	0,086	0,086	0,083	0,083	0,083
Splajny okresowe	0,768	0,770	0,769	0,590	0,591	0,591	1,814	2,251	2,687

1d, 2d, 3d oznaczają horyzonty prognozy odpowiednio na 1, 2 i 3 dni.

Źródło: Opracowanie własne.

## 5.7 Prognozowanie średniookresowe

Argumentu za prognozowaniem popytu na energię na 3 dni dostarcza Interdyscyplinarne Centrum Modelowania Uniwersytetu Warszawskiego, które publikuje prognozy na najbliższe 84 godziny<sup>47</sup> oraz Instytut Meteorologii i Gospodarki Wodnej, który w swoim serwisie przedstawia na bieżąco prognozy pogody na 5 kolejnych dni<sup>48</sup>. Ustaliłem 3 miesiące jako horyzont prognoz w moich analizach średniookresowych. Wydaje mi się, że ten okres jest dostatecznym okresem, żeby pokazać wpływ koniunktury na popyt na energię elektryczną.

<sup>46</sup>Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-Run Forecasts...*, str. 164.

<sup>47</sup><http://www.meteo.pl>

<sup>48</sup><http://www.pogodynka.pl/index>

Wykorzystanie informacji o warunkach atmosferycznych w prognozowaniu popytu na energię elektryczną jest analogiczne do przypadku krótkookresowego. Za prognozy pogody przyjąłem zaobserwowane wartości temperatury, zachmurzenia, nasłonecznienia i widoczności. Kwestia prognoz pogody dla trzech najbliższych miesięcy wydaje się być o wiele bardziej ważką przy prognozowaniu w średnim okresie. Horyzont 16-dniowy jest pokrywany predykcjami warunków atmosferycznych, natomiast rozwiązaniem dla pozostałych okresów prognozy może być użycie średnich historycznych wartości.

Oszacowałem wyprzedzający wskaźnik koniunktury (LEI) postaci (3.1)-(3.5) dla dłuższego zakresu danych od czerwca 1994 do listopada 2014<sup>49</sup>. Zgodnie z analizami bieżąca wartość tego wskaźnika powinna przybliżać koniunkturę jaka będzie miała miejsce za 3 miesiące. Praktyczne wykorzystanie tego indeksu polegać może na wyliczeniu jego wartości na danych zawierających bieżące obserwacje. W ten sposób dla ostatniego okresu w próbie  $T$  otrzymuje się prognozę zagregowanego stanu gospodarki dla okresu  $T + 3$ , oraz dla okresów  $T + 2$  i  $T + 1$ . Zmienne obserwowalne w modelu (3.1)-(3.5) to, zgodnie z tablicą 3.1 indeks giełdowy WIG20 i spread między rentownością 5-letnich obligacji skarbu państwa a 3 miesięcznym WIBORem. Wartości obydwu zmiennych są dostępne już po ostatnim dniu roboczym miesiąca, co sprawia, że nie napotykanym jest tutaj problem opóźnienia przy publikacji danych. Niektóre zmienne, jak na przykład indeks sprzedanej produkcji przemysłowej czy wskaźnik handlu detalicznego, dostępne są, zgodnie z harmonogramem Głównego Urzędu Statystycznego, dopiero po 35 dniach roboczych od zakończenia miesiąca. Wskaźnik LEI wyliczany ze zmiennych WIG20 i spreadu rentowności i oprocentowania jest wolny od tego problemu.

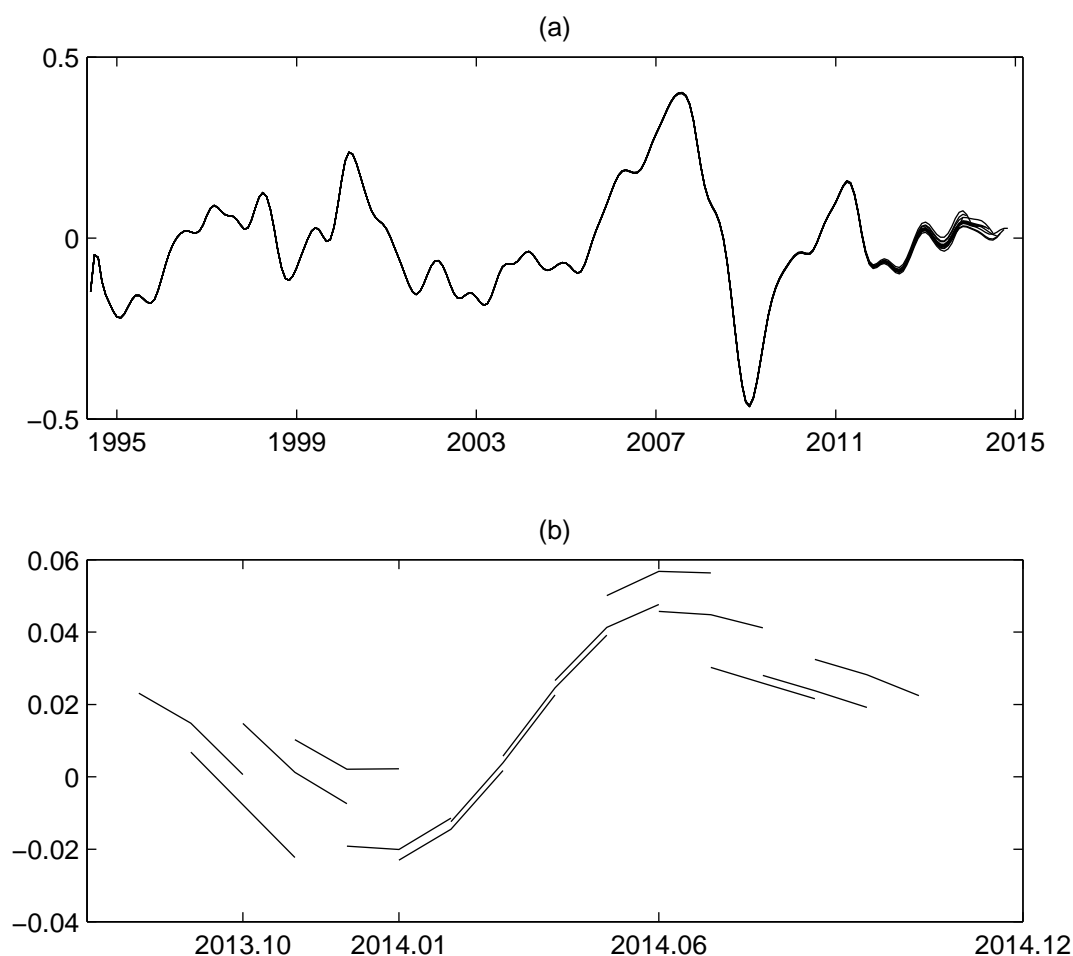
Za pomocą iteracyjnej procedury wyliczyłem wiązkę wartości wyprzedzającego wskaźnika koniunktury. W pierwszym kroku procedury wyznaczone zostały wartości LEI szacowane na próbie czerwiec 1994 - październik 2013 i w każdym kolejnym etapie zakres danych był rozszerzany o kolejny miesiąc. Powstała w ten sposób wiązka 14 wskaźników LEI obrazujących przyszły stan koniunktury, które zostały graficznie zaprezentowane na rysunku 5.1. Powodem, dla którego wyznaczyłem wiązkę wartości wskaźnika, był sposób estymacji. Wskaźnik koniunktury szacowany jest za pomocą filtru Kalmana, a wartości wyliczane są poprzez etapy filtrowania i wygładzania w filtrze Kalmana<sup>50</sup>. Poprzez wygładzanie uzyskuje się dokładniejsze oszacowania, gdyż do wyznaczenia wartości indeksu za okres  $t$ , gdzie  $1 < t < T$ , wykorzystywane są nie tylko obserwacje przeszłe  $1, \dots, t - 1$ , ale również  $t + 1, \dots, T$ . Na wygładzone wartości zmiennych stanu wpływ mają zatem wszystkie wartości obserwowanych zmiennych. Wpływ pojawienia się obserwacji za kolejny okres zaprezentowany jest na rysunku 5.1.

<sup>49</sup>Wyestymowałem ponadto równoczesny wskaźnik koniunktury i miesięczne PKB dla zbiorów danych powiększonych o obserwacje do grudnia 2014 roku. Z powodu zmiany metodologii wyliczania komponentów rachunków narodowych niedostępne wartości PKB za dwa ostatnie kwartały zastąpiłem własnym szacunkiem.

<sup>50</sup>Opis algorytmu wygładzania oraz filtru Kalmana znajduje się w aneksie.

Wyestymowałem ponadto równoczesny wskaźnik koniunktury i miesięczne PKB dla zbiorów danych powiększonych o obserwacje do grudnia 2014 roku.

Rysunek 5.1: Wiązka wskaźników wyprzedzających



Wykres (a) przedstawia wiązkę wskaźników LEI dla pełnego zakresu danych, natomiast wykres (b) 3 ostatnie wartości każdego z 14 szeregów w wiązce. *Źródło:* Opracowanie własne.

Do weryfikacji hipotezy o tym, że informacje o przyszłym stanie koniunktury pozwalają wyliczać dokładniejsze prognozy średniookresowe, zaproponowany został podobny eksperyment do opisanego w podrozdziale 5.1 i wykorzystanego do porównania modeli dla prognoz krótkookresowych. Powyższa hipoteza została sprawdzona przy wykorzystaniu modelu EGRV. Wybór modelu EGRV jest uargumentowany wynikami dla horyzontu 3 najbliższych dni, względnym nieskomplikowaniem oraz czasem estymacji. Dodatkowo, dla modelu EGRV można sprawdzić poprawność formy funkcyjnej oraz pozostałe założenia klasycznego modelu regresji liniowej.

Prawdziwość hipotezy o możliwości dokładniejszego prognozowania w średnim okresie

oznaczać musi, że dysponując informacjami o przyszłym stanie koniunktury można wyznaczyć średnio bliższe rzeczywistym wartościom prognozy. Analogicznie w przypadku, gdy najprostszy model niewykorzystujący szeregu przybliżającego koniunkturę pozwala uzyskiwać niegorsze wartości prognoz należałoby odrzucić postawioną hipotezę. Dzięki podziałowi zbioru danych na okres próby i poza próbą powyższą hipotezę można zweryfikować przy wykorzystaniu omawianych modeli i danych. Dla omówionych dalej modeli wyznaczone zostały prognozy na 1, 2 i 3 miesiące w przód, a okresem początkowym dla prognoz były w pierwszej iteracji okresy 30-dniowe po 10 października 2013 r. i w kolejnych 12 powtórzeniach okresy pełnych miesięcy zaczynając od 1 listopada 2013 roku.

Najprostsza wersja modelu EGRV została porównana ze swoimi czterema rozszerzeniami. Pierwszy z nich został rozszerzony o zmienną COIN, ilustrującą bieżący stan koniunktury, o której założono, że w momencie estymacji modelu w każdej iteracji znany jest stan koniunktury w 3 najbliższych miesiącach. Odpowiada to przypadkowi, w którym zawsze dostępne są informacje o cyklu koniunkturalnym w przyszłości. Druga wersja modelu EGRV, w tablicy 5.9 oznaczana jako „EGRV+MPKB”, porównywana z wersją bazową charakteryzowała się wykorzystaniem alternatywnej, do COIN, miary cyklu koniunkturalnego, którą był miesięczny produkt krajowy brutto, a w zasadzie jego komponent cykliczny. W trzeciej i czwartej wersji model EGRV rozszerzony został o wyprzedzający wskaźnik koniunktury w dwóch formach, jednej w oryginalnej postaci, a w drugiej wskaźnik został opóźniony dwa miesiące. Postępowanie to uzasadnione jest tym, że zaproponowany indeks LEI wyprzedza koniunkturę średnio o 3 miesiące. Dodatkowo, wyniki z rozdziału 3 wskazują, że zużycie energii elektrycznej poprzedza koniunkturę o 2 miesiące, więc wskaźnik przyszłego stanu gospodarki wyprzedza szereg energii tylko o 1 miesiąc. Hipoteza o tym, że informacje o stanie koniunktury pozwalają dokładniej prognozować popyt na energię elektryczną, musiałaby zostać odrzucona jeśli wykorzystanie szeregów koniunktury nie przełożyłoby się na wzrost dokładności prognoz z modelu EGRV.

Tablica 5.9: Błędy MAPE dla modeli EGRV dla prognoz średniookresowych w %

	Ogółem			Wiosna			Lato			Jesień			Zima		
	(1M)	(2M)	(3M)	(1M)	(2M)	(3M)	(1M)	(2M)	(3M)	(1M)	(2M)	(3M)	(1M)	(2M)	(3M)
<i>Dane oryginalne</i>															
Model EGRV	0,34	0,33	0,37	0,41	0,41	0,41	0,24	0,24	0,24	0,34	0,32	0,31	0,36	0,36	0,47
EGRV + COIN	0,32	0,32	0,35	0,38	0,38	0,38	0,20	0,20	0,20	0,34	0,32	0,35	0,36	0,36	0,46
EGRV + MPKB	0,33	0,33	0,36	0,37	0,37	0,37	0,20	0,20	0,20	0,36	0,35	0,38	0,36	0,36	0,46
EGRV + LEI3M	0,35	0,34	0,38	0,42	0,41	0,42	0,25	0,24	0,25	0,34	0,32	0,34	0,37	0,37	0,48
EGRV + LEI1M	0,34	-	-	0,42	-	-	0,24	-	-	0,33	-	-	0,37	-	-
<i>Dane bez dni specjalnych</i>															
Model EGRV	0,34	0,34	0,37	0,43	0,43	0,43	0,24	0,24	0,24	0,31	0,31	0,31	0,40	0,40	0,50
EGRV + COIN	0,32	0,32	0,36	0,39	0,39	0,39	0,21	0,21	0,21	0,31	0,31	0,31	0,39	0,39	0,50
EGRV + MPKB	0,32	0,32	0,36	0,37	0,37	0,37	0,20	0,20	0,20	0,33	0,33	0,34	0,37	0,37	0,48
EGRV + LEI3M	0,35	0,35	0,39	0,44	0,44	0,44	0,26	0,25	0,26	0,31	0,31	0,31	0,41	0,40	0,51
EGRV + LEI1M	0,38	-	-	0,48	-	-	0,32	-	-	0,30	-	-	0,43	-	-

1M, 2M, 3M oznaczają prognozy dla odpowiednio 1, 2 i 3 kolejnego miesiąca.

Źródło: Opracowanie własne.

Tablica 5.9 zawiera wyniki eksperymentu porównania różnych wersji modelu EGRV użytych do prognozowania średniookresowego. Przedstawione zostały w niej wartości średnich względnych błędów prognoz zarówno dla całego okresu out-of-sample, jak i jego podziału na pory roku. Kolumny (1), (2) i (3) oznaczają, że analizowanym horyzontem był odpowiednio pierwszy, drugi i trzeci kolejny miesiąc. Średnie wartości względnych błędów prognozy dla oryginalnych danych dają podstawę do nieodrżucenia postawionej hipotezy badawczej o wykorzystaniu cyklu koniunkturalnego w prognozowaniu średniookresowym. Z modelu EGRV rozszerzonego o równoczesny wskaźnik koniunktury lub o miesięczne PKB uzyskać można dokładniejsze wartości przewidywane niż z modelu bazowego, przy czym lepiej w roli dodatkowej zmiennej egzogenicznej sprawdza się ta pierwsza zmienna. Dla danych skorygowanych o dni specjalne wykorzystanie szeregu miesięcznego PKB jest najbardziej konkurencyjnym modelem. Różnica w błędach prognoz pomiędzy modelem EGRV ze zmienną COIN a bazowym modelem wynosi 0,0156 punktu procentowego dla prognoz jednomiesięcznych, 0,0138 p.proc. dla dwumiesięcznych i 0,0138 p.proc. dla trzymiesięcznych (w tablicy 5.9 różnice te w przybliżeniu wynoszą 0,02, 0,01 i 0,02 punktu procentowego). Wykorzystanie wskaźnika wyprzedzającego do przybliżenia stanu koniunktury w 3 najbliższych miesiącach pozwoliło poprawić jakość prognoz jedynie w horyzoncie drugiego miesiąca, natomiast wartości MAPE dla okresów pierwszego i trzeciego miesiąca są wyższe niż te uzyskiwane z modelu bazowego. Z przedstawionych wcześniej powodów model EGRV został także w jednej wersji rozszerzony o wskaźnik przyszłej koniunktury cofnięty o 2 miesiące (LEI1M). Wartości uzyskiwane z takiego modelu w horyzoncie najbliższego miesiąca są bliższe rzeczywistym o 0,0016 punktu procentowego niż prognozy z modelu bazowego.

Analiza jakości prognoz w podziale na pory roku wskazuje na największą poprawę w dla danych dla wiosny i lata, natomiast dla jesieni lepiej nie korzystać z szeregu ilustrującego koniunkturę. Błędy MAPE dla modeli EGRV ze zmiennym COIN i MPKB dla skorygowanych danych<sup>51</sup> niższe niż dla modelu bazowego z wyjątkiem prognoz dla trzeciego miesiąca w przód, którego wartości MAPE różnią się od bazowych wartości MAPE o 0,0029 punktu procentowego, a nie o 0,3 jak wcześniej. Bardzo ciekawego wniosku dostarcza analiza prognoz dla modelu EGRV rozszerzonego o opóźniony o 2 miesiące wskaźnik wyprzedzający, który choć charakteryzuje się istotnie wyższymi błędami dla całej próby out-of-sample, to jednak w okresie zimowym pozwala uzyskiwać najdokładniejsze prognozy ze wszystkich modeli.

Podsumowując stwierdzić należy, że druga główna hipoteza badawcza została zweryfikowana pozytywnie. W prognozowaniu średniookresowym, rozumianym jako horyzont 3-miesięczny, wykorzystanie informacji o stanie zagregowanej gospodarki pozwala wyznaczać dokładniejsze prognozy zapotrzebowania na energię elektryczną. Udowodniłem słuszność wykorzystania informacji o cyklu koniunkturalnym do wyliczania prognoz oraz zaproponowałem dwa rozwiązania, które mogą mieć zastosowanie w praktyce.

<sup>51</sup>Skorygowane zostały dni specjalne, czyli święta kościelne i państwowe.



# Zakończenie

Pierwsza główna hipoteza badawcza niniejszej pracy zweryfikowana została pozytywnie. Zarówno modele parametryczne jak i nieparametryczne wskazują na istnienie nieliniowej zależności między temperaturą a zapotrzebowaniem na energię. Literatura wskazuje na V-kształtność tej zależności w przypadku danych dla innych krajów i dla Polski otrzymałem bardzo zbliżoną postać, która obrazowo może być opisana jako przekrzywione L lub kij hokejowy. Choć w literaturze rzadko wykorzystywane są pozatemperaturowe pogodowe zmienne egzogeniczne, to w tym badaniu ustalona została istotność wpływu zachmurzenia, widoczności i nasłonecznienia. Niezwykle interesujący wniosek wyciągnąć można z analiz modeli regresji dla poszczególnych kwadransów, w których to nie temperatura, a zachmurzenie istotnie wydaje się różnicować zapotrzebowanie na energię elektryczną pomiędzy porami dnia. Bez wątplenia to właśnie dane wysokiej częstotliwości są idealne do analizowania relacji między warunkami atmosferycznymi a energią.

Potwierdziłem pozytywny wpływ informacji o koniunkturze na możliwości prognozowania w średnim okresie, będący meritum drugiej głównej hipotezy badawczej. Prognozy z modeli wykorzystujących zaproponowane indeksy koniunktury są dokładniejsze. Analizy wpływu cyklu koniunkturalnego na energię elektryczną wymagają jego liczbowych przybliżeń, więc zaproponowałem dwa równoczesne wskaźniki koniunktury. Analizę wpływu sytuacji gospodarczej na popyt na energię elektryczną poprzedziłem weryfikacją istnienia wahań o pożądanych częstościach w różnych szeregach energii elektrycznej.

Pozytywnie zweryfikowałem hipotezę poboczną o występowaniu potrójnej struktury sezonowej w obserwowanych danych wysokiej częstotliwości. Wartość dodaną pracy stanowi sformalizowana analiza wahań okresowych polskich danych. Uzyskane przeze mnie wyniki wskazują na stochastyczny, a nie deterministyczny charakter tych wahań. W związku z tym przedstawiłem sposoby modelowania danych ze złożoną stochastyczną strukturą sezonową. Wykorzystałem standardowe podejście oparte o zmienne zero-jedynkowe oraz użyłem trygonometrycznej postaci sezonowości.

Odrzuciłem dodatkową hipotezę badawczą o najniższym błędzie prognoz popytu na energię elektryczną z metody opartej o osobne modele regresji dla poszczególnych kwadransów. Otrzymałem wniosek, że jeden z najprostszych modeli, choć nie spełniał postulowanych w ekonometrii założeń, charakteryzował się najniższym błędem prognoz.

Moje badania szeregu zapotrzebowania na moce przesyłowe krajowego systemu elektroenergetycznego wskazały na obecność w nim niestacjonarnego komponentu wzrostowego. Analizę trendu wykonałem także w wersji uwzględniającej niestandardowe dni zmiany czasu. Na podstawie wyników nie mogłem odrzucić dodatkowej hipotezy badawczej o istnieniu trendu w analizowanych danych. Przedstawiłem także parametry filtru Hodricka-Prescotta dla przypadku danych 15-minutowych. Zaprezentowałem specjalną wersję testu na stacjonarność sezonowych danych wraz z tablicą wartości krytycznych.

Zaletą pracy jest to, że przedstawione w niej zostały różne podejścia do zagadnień napotkanych przy weryfikacji hipotez. W każdym takim przypadku zaprezentowałem klasyczne narzędzia oraz mniej standardowe i jednocześnie bardziej złożone metody. Testowanie hipotezy o istnieniu trendu w obserwowanych danych wykonałem dwoma sposobami. Wykorzystałem powszechnie stosowany rozszerzony test Dickey'a-Fullera oraz test Hylleberga i innych, który dopasowałem do danych wysokiej częstotliwości. Zastosowanie testu Hylleberga i innych uzasadniłem także możliwością sprawdzenia stochastycznego charakteru sezonowości. Dla większej odporności wyników wykorzystałem dwie nieparametryczne metody szacowania nieliniowej zależności między popytem na energię elektryczną a temperaturą. Pokazałem ponadto jak tą nieliniową zależność można modelować w liniowym modelu regresji. Za pomocą klasycznej analizy częstotliwości sprawdziłem istnienie wahań o częstościach z definicji cyklu koniunkturalnego, natomiast dla części interesujących szeregów o łamanych częstotliwościach zastosowałem wersję transformacji Fouriera dla danych nierówno rozłożonych w czasie. Wskaźnik równoczesny wyliczyłem z modelu zaproponowanego w literaturze, natomiast miesięczny komponent cykliczny PKB oszacowałem za pomocą niezaprezentowanego dotąd modelu.

Przedstawione analizy, schematy postępowania i niektóre chwytły mogą zostać zastosowane w analizie innych danych wysokiej częstotliwości. Budowane po kolei argumenty za potrzebą uwzględniania kolejnych komponentów szeregu czasowego w modelowaniu energii służą sporządzeniu optymalnego modelu wyjaśniającego i prognozującego zapotrzebowanie na moce przesyłowe krajowych sieci elektroenergetycznych. Prezentowane w każdej części pracy modele regresji liniowej służą zbudowaniu modelu jednocześnie dobrze wyjaśniającego zjawisko i umożliwiającego przetestowanie postawionych hipotez badawczych.

# Bibliografia

Abosedra S., Baghestani H., 1989, *New Evidence on the Causal Relationship between U.S. Energy, Consumption and Gross National Product*, Journal of Energy Development, 14, s. 285-292.

Adamowicz E., Dudek S., Pachnicki D., Walczyk K., 2009, *Synchronizacja cyklu koniunkturalnego polskiej gospodarki z krajami strefy euro w kontekście struktury tych gospodarek w Raport na temat pełnego uczestnictwa Rzeczypospolitej Polskiej w trzecim etapie Unii Gospodarczej i Walutowej*, Narodowy Bank Polski, Warszawa.

Akarca A.T., Long T.V., 1980, *On the Relationship between Energy and GNP: A Re-Examination*, Journal of Energy Development, 5, s. 326-331.

Al-Zayer J., Al-Ibrahim A.A., 1998, *Modeling the Impact of the Temperature on Electricity Consumption in Eastern Province of Saudi Arabia*, Journal of Forecasting, 15, s. 97-106.

Anderson R.C., Duru A., Reeb D.M., 2009, *Founders, Heirs, and Corporate Opacity in the United States*, Journal of Financial Economics, 92, s. 205-222.

ARE, *Informacja statystyczna o energii elektrycznej*, Agencja Rynku Energii, Warszawa.

ARE, *Statystyka Elektroenergetyki Polskiej*, Agencja Rynku Energii, Warszawa.

ARE, *Sytuacja energetyczna w Polsce - krajowy bilans energii*, Agencja Rynku Energii, Warszawa.

ARE, *Sytuacja w elektroenergetyce*, Agencja Rynku Energii, Warszawa.

Aries M.B.C., Newsham G.R., 2008, *Effect of Daylight Saving Time on Lighting Energy Use: A Literature Overview*, Energy Policy, 36, s. 1858-1866.

Artis M.J., Kontolenis Z.G., Osborn D.R., 1997, *Business Cycles for G7 and European Countries*, Journal of Business, 70(2), s. 249-279.

Aslan A., Apergis N., Yildirim S., 2014, *Causality between Energy Consumption and GDP in the U.S.: Evidence from Wavelet Analysis*, Frontiers in Energy, 8(1), s. 1-8.

Azevedo J.V.e, Koopman S.J., Rua A., 2003, *Tracking Growth and the Business Cycle: A Stochastic Common Cycle Model for the Euro Area*, Tinbergen Institute Discussion Paper, 2003-069/4.

Bigman D., Goldfarb D., Schechtman E., 1983, *The Journal of Futures Markets*, 3(3), s. 321-334.

Boot J.C.G., Feibes W., Lisman J.H.C., 1967, *Methods of Derivation of Quarterly Figures from Annual Data*, Journal of the Royal Statistical Society, Series C (Applied Statistics), 16(1), s. 65-75.

Bry G., Boschan Ch., 1971, *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, Columbia University Press.

Cancelo J.R., Espasa A., Grafe R., 2008, *Forecasting the Electricity Load from One Day to One Week Ahead for Short-Term Forecasting*, International Journal of Forecasting, 24, s. 603-615.

Charemza W., Deadman D. 1997, *Nowa ekonometria*, Polskie Wydawnictwo Ekonomiczne, s. 292.

Chauvet M., Piger J.M., 2005, *A Comparison of the Real-Time Performance of Business Cycle Dating Methods*, Working Papers 2005-021, Federal Reserve Bank of St. Louis.

Chow G.C., Lin A., 1971, *Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series*, The Review of Economics and Statistics, 53, s. 372-375.

Cochrane D., Orcutt G.H., 1949, *Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms*, Journal of the American Statistical Association, 44(245), s. 32-61.

Commandeur J.F., Koopman S.J., 2007, *An Introduction to State Space Analysis*, Oxford 2007.

Dickey D.A., Fuller W.A., 1979, *Distribution of Estimators for Autoregressive Time Series With a Unit Root*, Journal of American Statistical Association, Vol. 74, No. 366, s. 427-431.

Dickey D.A., Hasza D.P., Fuller W.A., 1984, *Testing For Unit Roots in Seasonal Time Series*, Journal of the American Statistical Association, 79, s. 355-367.

Di Fonzo T., 1990, *The Estimation of M Disaggregate Time Series when Contemporaneous and Temporal Aggregates Are Known*, Review of Economics and Statistics, s. 178-182.

Dordonnat V., Koopman S.J., Ooms M., Dessertaine A., 2008, *An Hourly Periodic State Space Model for Modeling French National Electricity Load*, International Journal of Forecasting, 24, 588-587.

Engle R.F., Granger C.W.J., Rice J., Weiss A., 1986, *Semiparametric Estimates of the Relation between Weather and Electricity*, Journal of the American Statistical Association, 81, s. 310-320.

Ferguson S.A., Preusser D.F., Lund A.K., Zador P.L., Ulmer R.G., 1995, *Daylight Saving Time and Motor Vehicle Crashes: The Reduction in Pedestrian and Vehicle Occupant Fatalities*, American Journal of Public Health, Vol. 85, No. 1, s. 92-95.

Forutna Z., Macukow B., Wąsowski J. (2001), *Metody numeryczne*, Wydawnictwa Naukowo Techniczne, Warszawa, s. 383.

Gardner E.S.J., McKenzie E., 1985, *Forecasting Trends in Time Series*, Management Science, 31, s. 1237-1246.

Gaudreault C., Lamy R., Liu Y., 2003, *New Coincident, Leading and Recession Indexes for the Canadian Economy: An Application of the Stock and Watson Methodology*, Department of Finance, Working Paper 2003-12.

Gradzewicz M., Growiec J., Hagemeyer J., Popowski P., 2010, *Cykl koniunkturalny w Polsce – wnioski z analizy spektralnej*, Bank i Kredyt, 41(5), s. 41-76.

Granger C.W.J., 1969, *Investigating Causal Relations by Econometric Models and Cross-Spectral Methods*, Econometrica, 37(3), s. 424-438.

Greene W.H., 2003, *Econometric Analysis 5ed*, Prentice Hall.

Greengard L.F., Lee J.-Y., 2004, *Accelerating the Nonuniform Fast Fourier Transform*, SIAM Review 46(3), 443-454.

GUS, 2014, *Równoległy oraz wyprzedzający zagregowany wskaźnik koniunktury, zegar koniunktury (Opracowanie eksperymentalne)*, Główny Urząd Statystyczny, Warszawa.

GUS, 2014, *Zużycie energii w gospodarstwach domowych w 2012 r.*, Główny Urząd Statystyczny, Warszawa.

Hamilton J.D., 1994, *Time Series Analysis*, Princeton University Press, Princeton.

Harding D., Pagan A., 2002, *Dissecting the Cycle: a Methodological Investigation*, Journal of Monetary Economics, 49, s. 365-381.

Hardle W.K., Muller M., Sperlich S., Werwatz A., 2004, *Nonparametric and Semiparametric Models, An Introduction*, Springer.

Harvey A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey A.C., Durbin J., 1986, *The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling*, Journal of the Royal Statistical Society series B, 48, s. 187-227.

Harvey A.C., Jaeger A., 1993, *Detrending, Stylized Facts and the Business Cycle*, Journal of Applied Econometrics, 8(3), s. 231-247.

- Harvey A.C., Koopman S.J., 1993, *Forecasting Hourly Electricity Demand Using Time-varying Splines*, Journal of the American Statistical Association, 88, s. 1228-1237.
- Harvey A.C., Koopman S.J., Penzer J., *Messy Time Series: A Unified Approach*, Advances in Econometrics, 13, s. 103-143.
- Harvey A.C., Koopman S.J., Riani M., 1997, *The Modeling and Seasonal Adjustment of Weekly Observations*, Journal of Business & Economic Statistics, Vol. 15(3), s. 354-368.
- Harvey A.C., Trimbur T.M., 2003, *General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series*, The Review of Economics and Statistics, 85(2), 244-255.
- Harvey A.C., Trimbur T., 2008, *Trend Estimation and the Hodrick-Prescott Filter*, Journal of the Japan Statistical Society, 38(1), s. 41-49.
- Henley A., Peirson J., 1997, *Non-Linearities in Electricity Demand and Temperature*, Oxford Bulletin of Economics and Statistics, 59.
- Hermans E., Wets G., Van den Bossche F., 2006, *Frequency and Severity of Belgian Road Traffic Accidents Studied by State-Space Methods*, Journal of Transportation and Statistics, 9(1), s. 63-76.
- Hill S.I., Desobry F., Garnsey E.W., Chong Y.-F., 2010, *The Impact on Energy Consumption of Daylight Saving Clock Changes*, Energy Policy 38, s. 4955-4965.
- Hodrick R.J., Prescott E.C., 1997, *Postwar U.S. Business Cycles: An Empirical Investigation*, Journal of Money, Credit and Banking, Vol. 29, No. 1, pp. 1-16.
- Holt C.C., 1957, *Forecasting Trends and Seasonal by Exponentially Weighted Averages*, Office of Naval Research Memorandum, 52. przedrukowane w Holt C.C., 2004, *Forecasting Trends and Seasonal by Exponentially Weighted Averages*, International Journal of Forecasting, 20 (1), s. 5-10.
- Hylleberg S., Engle R.F., Granger C.W.J., Yoo B.S., 1990, *Seasonal Integration and Cointegration*, Journal of Econometrics, Vol. 44(1-2), s. 215-238
- Hyndman R.J., De Livera A. M., Snyder R.D., 2010, *Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing*, Monash University, Working Paper 15/09
- Kalman R.E., 1960, *A New Approach to Linear Filtering and Prediction Problems*, Journal of Basic Engineering, Transactions of the ASME Series D, 82:35-45.
- Kim Ch.-J., Nelson Ch.R., 1999, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, Cambridge.
- Koopman S.J., Lee K.M., 2005, *Measuring Asymmetric Stochastic Cycle Components in U.S. Macroeconomic Time Series*, Tinbergen Institute Discussion Paper, 2005-081/4.

- Kotchen M.J., Grant L.E., 2008, *Does Daylight Saving Time Save Energy? Evidence from a Natural Experiment in Indiana*, NBER Working Paper 14429.
- Kraft J., Kraft A., 1978, *On the Relationship Between Energy and GNP*, Journal of the Energy Development, 3(2), s. 401-403.
- Lee C.-C., 2006, *The Causality Relationship between Energy Consumption and GDP in G-11 Countries Revisited*, Energy Policy, 34(9), s. 1086-1093.
- Lee C.-C., Chang C.-P., Chen P.-F., 2008, *Energy Consumption and GDP in Developing Countries: A Cointegrated Panel Analysis*, Energy Economics, 27(3), s. 415-427.
- Lee J.-Y., Greengard L.F., 2005, *The Type 3 Nonuniform FFT and its Applications*, Journal of Computational Physics, 206(1), 1-5.
- Lichota A., 2006, *Prognozowanie krótkoterminowe na lokalnym rynku energii elektrycznej*, Rozprawa doktorska, Akademia Górniczo-Hutnicza, Kraków, s. 176.
- Lisman J.H.C, Sandee J., 1964, *Derivation of Quarterly Figures from Annual Data*, Applied Statistics, 13, 87-90.
- Litterman R.B., 1983, *A Random Walk, Markov Model for Distribution of Time Series*, Journal of Business and Economics Statistics, 1, s. 169-173.
- Lorenc H. i inni (red.), 2012, *Kłęski żywiołowe a bezpieczeństwo wewnętrzne kraju*, Instytut Meteorologii i Gospodarki Wodnej, Państwowy Instytut Badawczy, Warszawa.
- Lokshin M., 2006, *Difference-Based Semiparametric Estimation in Partial Linear Models*, Stata Journal, Vol. 6(3), s. 377-383.
- Lutkepohl H., 2007, *New Introduction to Multiple Time Series Analysis*, Springer.
- Marcellino M., 1998, *Temporal Disaggregation, Missing Observations, Outliers, and Forecasting: A Unifying Non-Model-Based Procedure*, Advances in Econometrics, 13, s. 181-202.
- Mariano R.S., Murasawa Y., *A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series*, PIER Working Paper 02-014, Philadelphia 2002.
- Misiorek A., Weron R., 2004, *Modelowanie sezonowości a prognozowanie zapotrzebowania na energię elektryczną*, Energetyka 12/2004, s. 794-799.
- Mitchell W.C., 1927, *Business Cycles: The Problem and Its Setting*, National Bureau of Economic Research, Nowy Jork, s. 481.
- Narayan P.K., Narayan S., Smyth R., 2011, *Energy Consumption at Business Cycle Horizons: The Case of the United States*, Energy Economics, 33, s. 161-167.
- Nowicka-Zagrajek J., Weron R., 2002, *Modeling Electricity Loads in California: ARMA Models with Hyperbolic Noise*, Signal Processing 82, s. 1903-1915.

- Nunes L.C., 2005, *Nowcasting Quarterly GDP Growth in a Monthly Coincident Indicator Model*, Journal of Forecasting, 24, s. 575–592.
- Ozturk I., 2010, *A Literature Survey on Energy-Growth Nexus*, Energy Policy, 38(1), s. 340–349.
- Pardo A., Meneu V., Valor E., 2002, *Temperature and Seasonality Influences on Spanish Electricity Load*, Energy Economics, 24, s. 55–70.
- Parlament Europejski i Rada Unii Europejskiej, 2001, *Dyrektywa 2000/84/WE Parlamentu Europejskiego i Rady z dnia 19 stycznia 2001 r. w sprawie ustaleń dotyczących czasu letniego*, Dziennik Urzędowy Unii Europejskiej, 12 t.2.
- Pedregal D.C., Young P.C., 2006, *Modulated Cycles, an Approach to Modelling Periodic Components from Rapidly Sampled Data*, International Journal of Forecasting, 22 (1), s. 181–194.
- Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-run Forecast of Electricity Loads and Peaks*, International Journal of Forecasting, 13, s. 161–174.
- Rao N.K.J., 2003, *Small Area Estimation*, John Wiley & Sons, s. 344.
- Ravn M.O., Uhlig H., 2002, *On adjusting the Hodrick-Prescott filter for the frequency of observations*, The Review of Economics and Statistics, Vol. 84, No. 2, pp. 371–375.
- Robinson P.M., 1998, *Root-N-Consistent Semiparametric Regression*, Econometrica 56, s. 931–954.
- Rossi N., 1982, *A Note on the Estimation of Disaggregate Time Series When The Aggregate is Known*, The Review of Economics and Statistics, 64(4), s. 695–696.
- Sargan J. D., 1964, *Wages and Prices in the United Kingdom: A Study in Econometric Modelling*, w *Econometric Analysis for National Economic Planning*, ed. P.E. Hart, G. Mills, J.K. Whitaker, 25–64, London: Butterworths.
- Skrzypczyński P., 2008, *Wahania aktywności gospodarczej w Polsce i strefie euro*, Materiały i Studia nr 227, Narodowy Bank Polski.
- Skrzypczyński P., 2010, *Metody spektralne w analizie cyklu koniunkturalnego gospodarki Polskiej*, Materiały i Studia nr 252, Narodowy Bank Polski.
- Squalli J., 2007, *Electricity Consumption and Economic Growth: Bounds and Causality Analyses for OPEC members*, Energy Economics, 29(6), s. 1192–1205.
- Stern D.I., 1993, *Energy and Economic Growth in the USA: A Multivariate Approach*, Energy Economics, 15(2), s. 137–150.
- Stock J.H., Watson M.W., *A Probability Model of The Coincident Economic Indicators*,



Opublikowane w: G. Moore and K. Lahiri, editors. *The Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge 1990.

Szmit A., 2002, *Prognozowanie zapotrzebowania na energię elektryczną. Studium empiryczne dla regionu łódzkiego*, Rozprawa doktorska, Politechnika Łódzka, Łódź.

Taylor J.W., 2003, *Exponential Smoothing with a Damped Multiplicative Trend*, International Journal of Forecasting, 19, s. 715-725.

Taylor J.W., 2003, *Short-term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing*, Journal of Operational Research Society, 54, s. 799-805.

Taylor J.W., Buizza R., 2003, *Using Weather Ensemble Predictions in Electricity Demand Forecasting*, International Journal of Forecasting, 19, s. 57-70.

Taylor J.W., 2010, *Triple Seasonal Methods for Short-term Load Forecasting*, European Journal of Operational Research, 204, 139-152.

Tomaszewski M., 2005, *Model przedsiębiorstwa dystrybucyjnego działającego na otwartym rynku energii elektrycznej*, Rozprawa doktorska, Politechnika Opolska, Opole, s. 122.

Trimbur T.M., 2005, *Properties of Higher Order Stochastic Cycles*, Journal of Time Series Analysis, 27(1), s. 1-17.

U.S. Energy Information Administration, 2009, *Commercial Building Energy Consumption Survey*.

U.S. Energy Information Administration, 2009, *Residential Energy Consumption Survey*.

Valor E., Meneu V., Caselles V., 2001, *Daily Air Temperature and Electricity Load in Spain*, Journal of Applied Meteorology, 40, s. 1413-1421.

Verardi V., Debarsy N., 2012, *Robinson's Square Root-of-N-Consistent Semiparametric Regression Estimator in Stata*, Stata Journal, 12(4), s. 726-735.

Wei W.W.S., Stram D.O., 1990, *Disaggregation of Time Series Models*, Journal of the Royal Statistical Society. Series B (Methodological), 52(3), s. 453-467.

Weron R., 2006, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, John Wiley & Sons.

Weron R., 2014, *Electricity Price Forecasting: A Review of the State-of-the-Art with a Look into the Future*, International Journal of Forecasting, 30, s. 1030-1081

Weron R., Zator M., 2015, *A Note on Using the Hodrick-Prescott Filter in Electricity Markets*, Energy Economics, 48(C), s. 1-6.

Winters P.R., 1960, *Forecasting Sales by Exponentially Weighted Moving Averages*, Management Science, 6(3), s. 324-342.

Wolde-Rufael Y., 2006, *Electricity Consumption and Economic Growth: A Time Series Experience for 17 African Countries*, Energy Policy, 34(10), s. 1106-1114.

Woźniak R., 2011, *Metody rozszacowania kwartalnych indeksów koniunktury konsumenciej*, Wiadomości Statystyczne, 4, s. 1-10.

Yatchew A., 1997, *An Elementary Estimator of the Partial Linear Model*, Economic Letters 57, s. 135-143.

Yu E.S.H., Choi J.Y., 1985, *The Causal Relationship between Energy and GNP: An International Comparison*, Journal of Energy and Development, 10, s. 249-272.

Zani S., 1970, *Sui criteri di calcolo dei valori trimestrali di tendenza degli aggregate della contabilita nazionale*, Studi di Ricerche, Facolta di Economia e Commercio Universta degli Studi di Parma VII, s. 287-349.

# Spis tablic

1.1	Modele testujące stacjonarność i sezonowość danych . . . . .	15
1.2	Statystyki testowe i wartości krytyczne testu HEGY . . . . .	17
2.1	Oszacowania modelu Holta-Wintersa . . . . .	25
2.2	Testy diagnostyczne modelu Holta-Wintersa . . . . .	25
2.3	Oszacowania modelu Taylora . . . . .	27
2.4	Testy diagnostyczne modelu Taylora . . . . .	27
2.5	Testy diagnostyczne modelu Hyndmana i innych . . . . .	29
2.6	Wyniki dla modeli różnych sezonowości zero-jedynkowych . . . . .	37
2.7	Wyniki dla modeli różnych sezonowości trygonometrycznych . . . . .	40
3.1	Zmienne użyte do estymacji wskaźników koniunktury . . . . .	43
3.2	Oszacowania modelu wskaźnika równoczesnego . . . . .	46
3.3	Oszacowania modelu wskaźnika wyprzedzającego . . . . .	49
3.4	Oszacowania modelu miesięcznego PKB . . . . .	53
3.5	Porównanie punktów zwrotnych wskaźników i PKB . . . . .	55
3.6	Wyniki analizy korelacji krzyżowych komponentów cyklicznych . . . . .	57
3.7	Wyniki analizy korelacji krzyżowej między miesięcznymi szeregami energii elektrycznej a stanem gospodarki. . . . .	59
3.8	Punkty zwrotne szeregów produkcji i zużycia krajowego energii elektrycznej	61
3.9	Wyniki testu przyczynowości Grangera . . . . .	64
3.10	Wyniki estymacji modeli zawierających miesięczne PKB jako zmienną objaśniającą . . . . .	66
4.1	Fragment bazy danych pogodowych . . . . .	69
4.2	Podstawowe statystyki zmiennych pogodowych . . . . .	70
4.3	Miesięczne modele regresji . . . . .	72
4.4	Miesięczne modele regresji z danymi pogodowymi . . . . .	74
4.5	Modele regresji dla danych dziennych . . . . .	76
4.6	Dzienne modele z nieliniową zależnością między energią a temperaturą . .	81
4.7	Dni wolne od pracy i długie weekendy w latach 2002-2014 . . . . .	86
4.8	Dzienne modele z nieliniową zależnością między energią a temperaturą i czasem wolnym . . . . .	87

5.1	Daty przejścia orkanów w Polsce w latach 2002-2014 . . . . .	93
5.2	Błędy MAPE dla różnych wersji metody dnia podobnego w % . . . . .	96
5.3	Wybrane oszacowania modeli EGRV dla wybranych kwadransów . . . . .	100
5.4	Błędy MAPE dla modeli regresji w % . . . . .	103
5.5	Błędy MAPE dla modeli klasy ARMA . . . . .	105
5.6	Punkty węzłowe $\Delta$ w regresji splajnami kubicznymi . . . . .	112
5.7	Błędy MAPE dla modelu okresowych splajnów kubicznych w % . . . . .	113
5.8	Porównanie błędów MAPE dla modeli dla prognoz krótkookresowych w %	115
5.9	Błędy MAPE dla modeli EGRV dla prognoz średniookresowych w % . . .	119
5.10	Tablice wartości krytycznych testu HEGY dla poziomu istotności $\alpha = 5\%$ .	139
5.11	Porównanie modeli prognostycznych dla amerykańskich danych . . . . .	140

# Spis rysunków

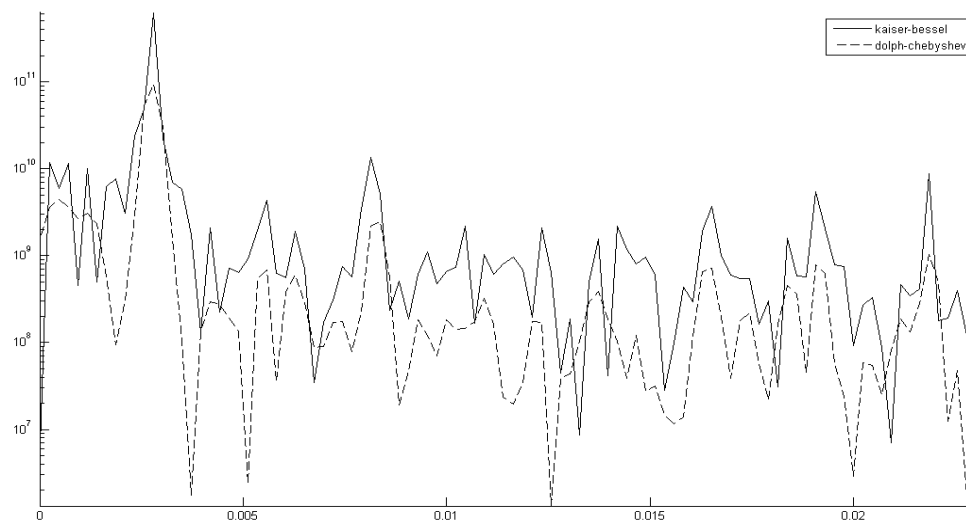
1.1	Zapotrzebowanie na moce przesyłowe KSE w trzecią środę miesiąca . . . .	9
1.2	Widmo oryginalnego szeregu - skala logarytmiczna . . . . .	11
1.3	Widmo szeregu z odjętym trendem liniowym - skala logarytmiczna . . . .	12
1.4	Widmo szeregu z odjętym trendem liniowym ograniczone do przedziału [0, 1] - skala logarytmiczna . . . . .	13
1.5	Widmo szeregu z odjętym trendem liniowym ograniczone do przedziału [0, 0,463] - skala logarytmiczna . . . . .	13
1.6	Część okienkowanego widma szeregu pozbawionego trendu deterministycz- nego w skali liniowej . . . . .	14
1.7	Szereg zapotrzebowania na moce przesyłowe i jego trend . . . . .	20
2.1	Funkcje wzmocnienia dla $(q_{\zeta}, q_{\omega}) \in \{(0,1^3, 0,5^2); (0,1^3, 0,5^4); (0,1^3, 0,5^6)\}$ . .	33
2.2	Funkcje wzmocnienia dla $(q_{\zeta}, q_{\omega}) \in \{(0,1^1, 0,1^1); (0,1^2, 0,1^2); (0,1^3, 0,1^3)\}$ . .	34
3.1	Logarytm krajowego zużycia energii elektrycznej i jego prognoza . . . . .	41
3.2	Komponent cykliczny równoczesnego wskaźnika koniunktury . . . . .	47
3.3	Komponent cykliczny wyprzedzającego wskaźnika koniunktury . . . . .	48
3.4	Komponenty cykliczne miesięcznego PKB i równoczesnego wskaźnika ko- niunktury . . . . .	53
3.5	Znormalizowane funkcje gęstości widmowej szeregów rocznych . . . . .	56
3.6	Funkcje gęstości spektralnych zmiennych miesięcznych . . . . .	58
3.7	Funkcje gęstości widmowej szeregów energii elektrycznej dostarczonej od- biorcom (a) ogółem, (b) na WN, (c) na SN i (d) na nN . . . . .	60
3.8	Zmienna objaśniająca skonstruowana z komponentu cyklicznego miesięcz- nego PKB . . . . .	65
4.1	Nieliniowy wpływ temperatury dla danych dziennych . . . . .	80
4.2	Wykres oszacowanych parametrycznych nieliniowych zależności . . . . .	82
4.3	Nieliniowy wpływ temperatury dla danych tygodniowych z 2012 roku . . .	83
5.1	Wiązka wskaźników wyprzedzających . . . . .	117
5.2	Widmo - skala logarytmiczna . . . . .	135



# Aneks

## Gęstość widmowa

Rysunek 5.2: Widmo - skala logarytmiczna



## Modele przestrzeni stanów

Modele przestrzeni stanów są modelami, które umożliwiają uwzględnienie zmiennych mających nieobserwowalny charakter. Generalnie modele przestrzeni stanów są matematycznymi modelami reprezentującymi fizyczny, dynamiczny system zmiennych wejściowych, wyjściowych i zmiennych stanu związanych równaniami różnicowymi pierwszego rzędu.

Założmy, że  $y_t$  oznacza  $(n \times 1)$  wektor zmiennych obserwowalnych w momencie czasu  $t$ ,  $\alpha_t - (r \times 1)$  nieobserwowalny wektor stanu,  $X_t - (n \times k)$  - wymiarowa macierz, której kolumnami są egzogeniczne lub ustalone zmienne. Macierz  $X_t$  może zawierać opóźnione wartości zmiennych pomiarowych  $y_t$  oraz te zmienne, które są nieskorelowane z  $\alpha_t$  oraz  $\varepsilon_t$ .

Wówczas system równań

$$\alpha_t = F\alpha_{t-1} + \zeta_t \quad (5.40)$$

$$y_t = A'X_t + H'\alpha_t + \varepsilon_t \quad (5.41)$$

opisuje dynamikę modelu. Macierze  $F$ ,  $A'$  oraz  $H'$  są macierzami o wymiarach odpowiednio  $(r \times r)$ ,  $(n \times k)$  oraz  $(n \times r)$ . Równanie (1) nazywane jest równaniem stanu (równaniem przejścia). Przedstawia ono dynamikę nieobserwowalnej zmiennej w modelu w reakcji na zmiany opóźnionych zmiennych zależnych oraz wektora  $\zeta_t$  zaburzeń losowych. Równanie (2) nazywane jest równaniem pomiarowym (równaniem sygnałowym), odzwierciedla zależność między wektorem zmiennych pomiarowych (obserwowalnych) a nieobserwowalnym wektorem stanu i zmiennymi egzogenicznymi  $X_t$ . W równaniu tym pojawia się dodatkowo  $\varepsilon_t$  -  $(n \times 1)$ -elementowy wektor, reprezentujący wahania nie tłumaczone poprzez zmienne nieobserwowalne i egzogeniczne. Wektory błędów  $(r \times 1)$ -elementowy  $\zeta_t$  oraz  $(n \times 1)$ -elementowy  $\varepsilon_t$  są białymi szumami, są niezależne i charakteryzują się wariancją stałą w czasie.

$$\mathbb{E}(\zeta_t \zeta_\tau) = \begin{cases} Q & \text{dla } t = \tau \\ 0 & \text{w p.p.} \end{cases}$$

$$\mathbb{E}(\varepsilon_t \varepsilon_\tau) = \begin{cases} R & \text{dla } t = \tau \\ 0 & \text{w p.p.} \end{cases}$$

$$\mathbb{E}(\varepsilon_t \zeta_\tau) = 0 \text{ dla } \text{każdych } t \text{ i } \tau$$

gdzie  $Q$  i  $R$  są macierzami o wymiarach odpowiednio  $(r \times r)$  i  $(n \times n)$ . Dodatkowo potrzebne jest założenie o wektorze stanu  $\alpha_1$ , o którym zakłada się, że jest nieskorelowany z realizacjami  $\varepsilon_t$  oraz  $\zeta_t$ , czyli  $\mathbb{E}(\zeta_1 \alpha_1) = 0$  oraz  $\mathbb{E}(\varepsilon_1 \alpha_1) = 0$  dla  $t = 1, \dots, T$ .

Reprezentacja modelu przestrzeni stanów umożliwia zastosowanie filtru Kalmana jako algorytmu estymującego i prognozującego. Filtr Kalmana umożliwia wyznaczenie prognoz na jeden okres w przód o minimalnym błędzie średniokwadratowym (MSE). Dodatkowo, jeżeli model ma liniową postać i błędy mają rozkład normalny, to uzyskiwane prognozy są estymatorami MNW.



## Filtr Kalmana

Filtr Kalmana został zaproponowany przez

Po zapisaniu modelu w postaci przestrzeni stanów filtr Kalmana może zostać użyty do estymacji parametrów. Filtr składa się z  $T$  kroków, z który każdy zawiera etap predykcji i korekty. Etap predykcji, dla  $1 \leq t \leq T$ , składa się z czterech równań

$$\alpha_{t|t-1} = F\alpha_{t-1|t-1}, \quad (5.42)$$

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q, \quad (5.43)$$

$$v_{t|t-1} = y_t - H\alpha_{t|t-1}, \quad (5.44)$$

$$f_{t|t-1} = HP_{t|t-1}H' + R, \quad (5.45)$$

gdzie  $\alpha_{t|s} = \mathbb{E}(\alpha_t|y_1, \dots, y_s)$  oraz  $P_{t|s} = \text{Cov}(\alpha_t|y_1, \dots, y_s)$ . Etap korekty,  $1 \leq t \leq T$ , składa się z dwóch równań

$$\alpha_{t|t} = \alpha_{t|t-1} + P_{t|t-1}H'f_{t|t-1}^{-1}v_{t|t-1}, \quad (5.46)$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H'f_{t|t-1}^{-1}HP_{t|t-1}. \quad (5.47)$$

Filtr Kalmana pozwala wyznaczyć  $\alpha_{t|t}$ , czyli wartości zmiennych stanu na podstawie próbki  $y_1, \dots, y_t$ . Wartości  $\alpha_{t|t-1}$  nazywane są prognozowanymi, natomiast  $\alpha_{t|t}$  odfiltrowanymi. Przedmiotem zainteresowania mogą być też wartości zmiennych stanu wyznaczone na podstawie informacji z całej próby, to jest  $y_1, \dots, y_T$ . Algorytm rekursji wstecznej, nazywany wygładzaniem, pozwala wyznaczyć  $\alpha_{t|T}$  dla  $t < T$ . Pojedynczy krok rekursji wstecznej dla  $t < T$  ma postać

$$P_t^* = P_{t|t}F_{t+1}'P_{t+1|t}^{-1} \quad (5.48)$$

$$\alpha_{t|T} = \alpha_{t|t} + P_t^*(\alpha_{t+1|T} - F_{t+1}\alpha_{t|t}) \quad (5.49)$$

$$P_{t|T} = P_{t|t} + P_t^*(P_{t+1|T} - P_{t+1|t})P_t^{*'} \quad (5.50)$$

gdzie  $\alpha_{T|T} = \alpha_{T|T}$  oraz  $P_{T|T} = P_{T|T}$ . Szereg wartości  $\alpha_{t|T}$  nazywany jest szeregiem wygładzonym.

Opisy filtru warte polecenia zawarte są w książkach Kima i Nelsona (1999) oraz Lutkepohl'a (2007).

## Testy diagnostyczne

Najważniejszym założeniem na temat reszt jest normalność. Sprawdzona być może za pomocą statystyki Ljunga-Boxa, która jest zdefiniowana dla opóźnień  $l = 1, 2, \dots, k$  w następujący sposób:

$$Q(k) = n(n+2) \sum_{l=1}^k \frac{r_l^2}{n-l},$$

gdzie  $n$  jest liczbą obserwacji,  $r_k$  oznacza autokorelację reszt dla opóźnienia  $k$  i jest zdefiniowana jako

$$r_k = \frac{\sum_{i=1}^{n-k} (e_t - \bar{e})(e_{t+k} - \bar{e})}{\sum_{t=1}^n (e_t - \bar{e})^2},$$

gdzie  $\bar{e}$  jest średnią z  $n$  reszt. Statystyka testowa ma rozkład  $\chi^2$  z  $(k - w + 1)$  stopniami swobody, gdzie  $w$  jest liczbą estymowanych hiperparametrów, czyli wariancji zaburzeń. Według hipotezy zerowej nie występuje autokorelacja dla opóźnień  $l = 1, 2, \dots, k$ .

Drugim, pod względem ważności, założeniem jest homoskedastyczność reszt. Stałość wariancji testuje się za pomocą testu Goldfelda-Quandta, w którym statystyka testowa ma postać

$$H(h) = \frac{\sum_{t=n-h+1}^n e_t^2}{\sum_{t=d+1}^{d+h} e_t^2},$$

gdzie  $d$  jest liczbą estymowanych parametrów za pomocą rozmytej inicjalizacji (*ang.* diffuse initialisation), natomiast oznacza najbliższą liczbę całkowitą do wartości  $(n - d)/3$ . Statystyka testowa ma rozkład  $F$  z  $(h, h)$  stopniami swobody. W przypadku gdy  $H(h)$  przyjmuje wartość większą od 1 wystarczy sprawdzić czy zachodzi  $H(h) < F(h, h)$ . W przeciwnym przypadku, gdy  $H(h)$  jest mniejsze od 1, należy sprawdzić czy  $1/H(h) < F(h, h)$ . Zgodnie z hipotezą zerową występuje homoskedastyczność reszt.

Ostatnie założenie dotyczy normalności rozkładu reszt. Testuje się ją za pomocą testu Jarque-Bera. Statystyka testowa jest zdefiniowana w sposób:

$$N = n \left( \frac{S^2}{6} + \frac{(K - 3)^2}{24} \right),$$

gdzie  $S$  oznacza skośność, natomiast  $K$  - kurtozę reszt. Statystyka ma rozkład  $\chi^2$  z dwoma stopniami swobody. Hipoteza zerowa w tym teście stwierdza normalność rozkładu reszt.

## Tablice testu HEGY

Tablica 5.10: Tablice wartości krytycznych testu HEGY dla poziomu istotności  $\alpha = 5\%$ .

N	t_stat	F_stat1	F_stat2
500	-3,15	2,32	3,10
1 000	-2,95	2,92	3,29
1 500	-2,89	3,13	3,34
2 000	-2,94	3,22	3,36
2 500	-2,87	3,33	3,37
3 000	-2,85	3,20	3,39
3 500	-2,87	3,30	3,39
4 000	-2,86	3,33	3,38
4 500	-2,88	3,32	3,39
5 000	-2,88	3,32	3,40
5 500	-2,88	3,30	3,40
6 000	-2,88	3,35	3,40
6 500	-2,90	3,43	3,40
7 000	-2,89	3,39	3,41
7 500	-2,90	3,43	3,40
8 000	-2,86	3,38	3,40
8 500	-2,86	3,43	3,41
9 000	-2,86	3,49	3,41
9 500	-2,88	3,41	3,41
10 000	-2,86	3,34	3,41
412 892	-2,82	3,37	3,43

Statystyka  $t$  używana jest przy testowaniu hipotezy o zintegrowaniu szeregu. Za pomocą statystyk  $F\_stat1$  i  $F\_stat2$  zweryfikować można istnienie odpowiednio stochastycznej oraz deterministycznej sezonowości.

## Wyniki zawodów prognostycznych z lat 1990-1992

Tablica 5.11: Porównanie modeli prognostycznych dla amerykańskich danych

Godzina	Pon	Wt	Śro	Czw	Pt	Sob	Nie
01:00	B	A	A	A	FS	SHELF	SHELF
02:00	STAT	STAT	A	A	FS	L	STAT
03:00	FS	STAT	B	SHELF	FS	A	FS
04:00	FS	B	B	B	FS	A	SHELF
05:00	STAT	B	B	B	SHELF	SHELF	SHELF
06:00	B	FS	B	B	FS	B	B
07:00	A	FS	B	B	FS	FS	B
08:00	A	FS	A	B	FS	FS	B
09:00	STAT	FS	A	A	FS	B	B
10:00	STAT	FS	A	A	B	A	STAT
11:00	FS	FS	A	A	A	B	L
12:00	FS	B	A	A	A	L	L
13:00	L	B	A	A	A	L	L
14:00	B	FS	B	B	B	L	B
15:00	B	B	B	B	A	L	L
16:00	A	A	B	B	B	L	A
17:00	A	B	B	B	B	L	L
18:00	L	A	B	B	B	L	FS
19:00	FS	A	A	A	A	FS	L
20:00	L	B	A	A	A	L	FS
21:00	FS	B	A	B	A	L	L
22:00	FS	B	B	B	A	L	L
23:00	FS	A	A	B	L	SHELF	ANN2
24:00	FS	L	B	A	A	SHELF	B
Szczyt poranny	A	FS	B	B	FS	B	B
Szczyt wieczorny	L	A	B	B	B	L	FS

Oznaczenia: A i B oznaczają model zaproponowany przez Ramanathana i innych<sup>52</sup> w wersji podstawowej i rozszerzonej, SHELF - model zmiennych w czasie okresowych splajnów kubicznych Harvey'a i Koopmana<sup>53</sup>, L - model spółki Puget Sound Power and Light, natomiast Ramanathan i inni<sup>54</sup> odnotowali brak informacji o modelach FS i ANN2.

Źródło: Ramanathan R., Engle R., Granger C.J.V., Vahid-Araghi F., Brace C., 1997, *Short-run Forecast...*, str. 169.